SUPPLEMENTAL MATERIAL

SUPPLEMENTAL METHODS

All data, analytic methods and study materials are available from the corresponding author upon request.

Animal models

All animal requisitions, housing, treatments and procedures were performed according to all state and institutional laws, guidelines and regulations. All studies were approved by the Ethics Committee for Animal Research at the University of Navarra and the Government of Navarra. *Collagen1α1-GFP* and *Cthrc1-KO* mice were previously described[15, 22].

Induction of MI in mice

Myocardial infarction (MI) was induced in mice by ligation of the left anterior descending (LAD) coronary artery as previously described[5]. Briefly, 8 to 10 weeks old mice were anesthetized with vaporized isoflurane, intubated using a 20G intravenous catheter, mechanically ventilated, and placed on a heating pad to maintain body temperature. A left thoracotomy was performed at the fourth-fifth intercostal space, where muscles were dissected. The LAD coronary artery was permanently ligated using a 7/0 non-absorbable ethylene suture. After visual verification of anemia and akinesis of the apex and anterior-lateral wall to ensure coronary occlusion, the thorax was closed in layers. After extubation, mice were kept warm until fully recovered. Healthy or infarcted mice at 3, 5, 7, 14 or 30 dpi were sacrificed and processed for analyses.

Model of chronic cardiac fibrosis by Angiotensin-II infusion

Mini-osmotic pumps (Alzet, models 1002, 1007D, and 1004) were implanted subcutaneously on the back of female and male 129SvEv strain (Taconic, Germantown, NY) mice at 8 to 10 weeks of age. Pumps were primed in PBS overnight at 37 °C and delivered angiotensin-II (Bachem, H-1705) at 0.7 mg/kg/day or vehicle (PBS) for up to 28 days. Hearts were harvested for analyses at 2, 7, 14 and 28 days of angiotensin-II treatment.

Induction of MI in pig

Experimental myocardial infarction was performed in approximately 12-15 month-old young adult male Göttingen minipigs weighing 20-30 kg (Ellegaard; Soroe Landevej 3024261, Dalmose, Denmark) (n=21). Pigs were premedicated with Amiodarone (400 mg/day for the first week and 200 mg/day for the second week) for 14 days prior to infarct induction. On D-1 Aspirin (300 mg) and Clopidrogel (300 mg) was administered. For the infarct procedure, animals were weighed and sedated with a mixture of tiletamine and zolepam (Zoletil®) 5 mg/kg i.m. Following adequate sedation, the neck was shaved, properly scrubbed and disinfected with Povidone Iodure and alcohol, followed by transfer to the surgery room. One intravenous catheter was placed in a marginal vein of one ear. An appropriately sized endotracheal tube was inserted for mechanical ventilation. Body temperature was maintained at 36.5-39 °C during the procedure with a blanket. Surface electrocardiogram (ECG) (lead II) was placed and connected to a DASH 5000 to monitor the onset of arrhythmias (ventricular tachycardia and fibrillation). To induce a stable and prolonged anesthesia, an i.v. propofol infusion was initiated once the pig was under ventilation (30% O_2, ~400 mL tidal volume, ~18 breaths per min). Propofol dosing was adjusted to ~10 mg/kg per hour throughout the experimental procedure, depending on the anesthetic state of each animal.

Adequate anesthesia was checked continuously. The animal was fitted with sterile drapes and sterile instruments were used by the surgeon. All surgical procedures were performed using standard sterile techniques, including sterile gowns, gloves, masks and instruments. Remifentanil 18-20 μg/kg/hr IV was given for intraoperative analgesia.

A small incision was performed in the neck and, using the Seldinger technique and a trocar was inserted into the carotid artery. Under fluoroscopy guidance, a JR 3.5 catheter was advanced over the wire to the level of the coronary sinus and placed at the coronary ostium without full engagement. Good placement was confirmed by angiography with a bolus of contrast agent (VISIPAQUE) to visualize the left main, circumflex and anterior descending coronary arteries. Heparin (300 UI/kg i.v.) was given to prevent clotting just before LAD coronary artery occlusion. A guide wire was advanced into the LAD coronary artery and a balloon catheter (3.5 x 15-20) was advanced just below the first diagonal and inflated at 4-6 atm (replace with SI unit Pascal) to ensure occlusion for 150 min of the LAD coronary artery confirmed by coronary angiogram. The balloon catheter was then deflated, slowly removed and reperfusion documented by coronary angiography. In case of ventricular fibrillation, defibrillation was immediately performed using an HP defibrillator. A bolus of lidocaine (2 mg/kg i.v.) and nitroglycerin (40 μg/kg intracoronary) was administered to avoid vasospasms and arrhythmia. During the recovery period, phentanyl patches (0.03 mg/kg transdermal) were administered for analgesia. Animals were allowed to recover from anesthesia under a heat lamp in individual cages and extubation was performed when the pig began to chew. Amoxicillin (Dufamox® 0.015 mg/kg i.m. antibiotic) and Carprofen (2 mg/kg s.c. NSAIDS) was given for post-procedural care.

At 8 days, 2 and 6 months post-ischemia, animals were sacrificed with pentobarbital and a saturated solution of potassium chloride. Hearts were excised and cut in 6 transversal sections **(Figure 6A)**. Half of the slices were fixed in Zn-Formalin, paraffin embedded and sectioned for histological assessment. The other half was dissected into four representative blocks (1-2 cm^2) according to their anatomical region of infarction (Infarct (IZ), border (BZ), and remote zones (RZ)) followed by snap freezing for molecular analyses. Two control animals were included in the study.

Heart samples from pigs with a subacute infarct were kindly provided by the laboratory of Professor S. Janssens (Univ. of Leuven). Adult domestic farm pigs (body weight: 30-34 kg) were used for the study. Briefly, under fluoroscopic guidance, a 7-French catheter was inserted via the left coronary ostium just below the LAD distal to the first diagonal branch and the balloon catheter was inflated producing the MI. Occlusion was maintained for 75 min. After deflation, restoration of LAD coronary artery blood flow was verified by coronary angiography and ST segment changes by ECG. Eight or sixty days after MI induction, animals were euthanized and heart biopsies from IZ and RZ of the heart were collected for molecular and histological assessment.

Cardiac function evaluation in pig

The week before induction of MI and 180dpi (end of study) 21 animals were evaluated by echocardiography and magnetic resonance. Briefly, following adequate sedation with Zoletil® and induction of a stable and prolonged anesthesia, isoflurane inhalator, the left side of the chest was shaved and properly cleaned. Echocardiographic examination was carried out using a cardiac ultrasound system equipped with a 4 MHz transducer (Philips SONOS 5500 M 2424A). The echocardiographic images were obtained using an approach from the right (4th or 5th intercostal

space, in left decubitus) for parasternal long-axis and short-axis views. Three consecutive acquisitions for each view were performed and analyzed by two independent trained readers. Left ventricular volumes were calculated using the Simpson's method. Stroke volume was determined as: SV (mL) = LV diastolic volume – LV systolic volume. The ejection fraction was determined as EF (%) = (SV / LV diastolic volume) x 100 and cardiac output as CO (mL/min) = SV x heart rate. Samples were classified by preserved (>45%) or reduced (≤45%) EF[40] ($p \leq 0.0001$, Two-tailed *t*-test with a Mann-Whitney post-hoc test).

Cardiac magnetic resonance examinations were performed on a 1.5 Tesla system (MAGNETOM Symphony with Tim, Siemens Healthcare GmbH, Erlangen, Germany) equipped with commercially available cardiac MRI software, electrocardiographic triggering and cardiac-dedicated surface coils. Pigs were placed in the left lateral decubitus position. Images were acquired with mechanical ventilation. MRI-compatible ECG pads were placed on the chest, using the guidelines of the manufacturer, and connected to the scanner. The ECG waveforms were examined for a clear R wave and adjusted if necessary. Standard steady-state free-precession (SSFP) cine images were acquired in the four-chamber, two-chamber and short-axis views to assess left-ventricular function with the following parameters: 8 mm slice thickness, no gap, TR= 43.26 ms, TE= 1.3 ms, flip angle= 80, 156 x 192 matrix, field of view of 260-280 x 325-375 mm, 1.7 x 1.7 mm pixel size, 14 segments, and 25 phases. To assess myocardial area at risk, a T2-STIR sequence was used in the short axis view. Rest perfusion was performed after the injection of 0.05 mmol/kg of intravenous contrast (gadobutrol, Gadovist, Bayer Schering Pharma AG, Berlin-Wedding, Germany) using an automated dual-head injector (Medrad Inc, Warrendale, Pennsylvania, USA) at 4 mL/s. An additional dose of gadolinium (0.1 mmol/kg) was administered

(total dose of 0.15 mmol/kg) before acquiring early and delayed gadolinium enhanced images, for which a turbo-flash phase-sensitive inversion recovery (PSIR) sequence (FA= 25°, TR/TE= 53.68/3.19 ms, voxel size= 1.4 × 1.4 × 8 mm) was employed. Sequence parameters were modified when necessary. Samples were classified as preserved (>45%) (n=9) or reduced (≤45%) (n=12) EF ($p \leq 0.0001$, two-tailed *t*-test with a Mann-Whitney post-hoc test). The infarct area was determined by resonance imaging ($p \leq 0.05$, Two-sided *t*-test).

Human samples

The study protocol was approved by the Medical Ethics Committee and informed consent was obtained from all patients. For RNA studies, human heart samples were kindly provided by Professor S. Janssens. Heart biopsies from the right (RV) and left ventricle (LV) of healthy donors (n=6), from patients with dilated cardiomyopathy (DCM) (n=5), and from infarct (IZ) and remote zone (RZ) of patients with a history of chronic MI (n=8) were collected for molecular analysis. For histological analysis, human heart samples (n=2) were kindly provided by the Maine Medical Center BioBank.

Single cell optimization and preparation

Mouse cardiac interstitial cells (CIC) were obtained from individual, 8-10 week-old mice as previously described[5]. Briefly, after euthanasia the thorax was opened and the heart was perfused with ice-cold phosphate buffered saline pH 7.6 (PBS) (Lonza), atria were excised and discarded, and both ventricles were digested completely. When appropriate hearts were divided into remote, border and infarct zone (i.e. for zonal RNA-seq). Excised tissues were placed in ice cold DMEM medium (Sigma) supplemented with 10% Fetal Bovine Serum (FBS) (Hyclone, GE). Ventricles were minced using a sterile scalpel. Pieces of tissue were incubated on an orbital shaker for 10 min

at 37 °C in the presence of Liberase TH (125 µg/mL) (Roche) in HBSS++ solution (Hanks balanced salt solution, Gibco). After enzymatic incubation, partially digested tissue was mechanically dissociated by slowly pipetting to generate a single cell suspension. The supernatant was filtered through a cell strainer to discard cardiomyocytes (CM) (40 µm, nylon; Falcon). The digestion was repeated with the sedimented pieces and the supernatants were pooled together. Erythrocytes were removed using RBC lysis buffer (eBioscience). The total time for enzymatic digestion was 30 min.

In order to include all cardiac fibroblasts (CF) in the following studies, an optimization step was performed. Total CIC pellet was resuspended in 80 µL sorting buffer (2 mM EDTA, 0.5% BSA in PBS) and incubated with 20 µL of Feeder Removal MicroBeads (mEFSK4) (Miltenyi) for 15 min at 4 °C. A positive selection of CIC, enriched in CF, was performed twice using LS columns (Miltenyi) according to the manufacturer's instructions. However, in order to check if the selection was optimal, both fractions, negative (fraction 1, pass through column) and positive (fraction 2, bind to column) were sorted by GFP fluorescence **(Figure IE in the Supplement)**. No GFP^+ cells were found in the negative fraction. Both fractions were resuspended in complete medium (DMEM + 10% FBS + 1% P/S (Life Technology) + 1% L-Glutamine (Life Technology)) containing 10 ng/mL of basic fibroblasts growth factor (bFGF/FGF2, Peprotech) and plated on 0.1% (w/v) gelatin-coated wells for 4 days (Sigma). Only GFP^+ cells sorted from fraction 2 (named "fraction 3" in **Figure IE in the Supplement**) contained CF (documented in **Figure IF in the Supplement**) were used in all the following experiments as $GFP^+/CD31^-/CD45^-$.

Flow cytometry analysis

For flow cytometry characterization of *Col1α1-GFP*$^{+}$ CIC cells, the pellet was resuspended in 100 µL of sorting buffer before staining with the respective antibodies for flow cytometry or fluorescence activated sorting **(Table I in the Supplement)**. After 15 min incubation at room temperature in the dark, calcein-violet (Life Technologies) and/or Vybrant® DyeCycle™ Orange (VDO, Life Technologies) were added to antibody/cell suspensions at final concentrations of 1 and 1.25µM respectively. After incubation with dyes in a 37 °C water bath for 10 min, samples were washed twice with sorting buffer and spun at 1,200 rpm for 5 min. Then supernatant was discarded, and the final pellet was resuspended in 250 µL of sorting buffer. Just before flow cytometry, TO-PRO-3 (Invitrogen) was added to each sample to determine viability. Events were gated for metabolically active (calcein^{+}), nucleated (VDO^{+}), and viable (TO-PRO-3^{-}) single cells **(Figure 1B)**, before gating on CD31^{-}/CD45^{-} events for the analysis of GFP^{+} CF, CD31^{+}/CD45^{-} for endothelial cells, and CD45^{+}/CD31^{-} for bone marrow-derived cells **(Figure 1B)**. Flow cytometry for these samples were performed on a FACSCANTO Flow Cytometer (BD Biosciences). The files generated by flow cytometry were processed using FlowJo Software (Tree Star, Ashland, USA). Statistical significance was analyzed by Student's *t*-test and shown as mean ± SD using GraphPad Prim 6.0 (GraphPad Software) ($p<0.05$). Two to five biological replicates were examined for healthy mice, and 3 to 5 for each time point for infarcted ones.

For flow cytometry characterization of *Cthrc1-KO* and *Cthrc1-WT* CIC cells, individual pellets were resuspended in 100 µL of sorting buffer and treated with Human TruStain fcX™ (anti-mouse 16/32 antibody, BioLegend, San Diego, CA) to prevent non-specific binding, followed by incubation with relevant antibodies for 25 min at 4 °C. Cell-surface antigen expression was

examined for $CD31^+$ and $CD45^+$ cells. For intracellular CTHRC1 staining, cell suspensions were incubated in DMEM containing 5% FBS and 3 μg/mL Brefeldin A (BioLegend, San Diego, CA) for 5 hours. After treatment with a Cytofix/Cytoperm kit (BD Biosciences), the permeabilized cells were stained with polyclonal anti-CTHRC1 antibody[13] in combination with PE-conjugated anti-rabbit IgG (Jackson ImmunoResearch Laboratories, Inc). Viable and non-viable cells were distinguished using LIVE/DEAD® Fixable Violet Stain kit (Life Technologies). Flow cytometry for these samples was conducted on a MacsQuant Analyzer 10 (Miltenyi Biotec., Inc.) and the data was processed using WinList 5.0 software (Verity software House). Three biological replicates per condition were analyzed.

Cell sorting

The positive fraction from the LS column enrich in CF was centrifuged at 1,200 rpm and the pellet was resuspended in 100 μL of sorting buffer. Cells were incubated with the corresponding antibodies for 15 min at room temperature in the dark **(Table I in the Supplement)**. After incubation, the samples were washed twice with sorting buffer and spun at 1,200 rpm for 5 min, supernatant was discarded, and the final pellet was resuspended in 250 μL sorting buffer. Cell sorting was performed using a FACSAria (BD Biosciences) and analyzed with FACSdiva software (BD Biosciences). Standard, strict forward scatter width *versus* area criteria were used to discriminate doublets and gate only singleton cells. Viable cells were identified by staining with 7-AAD (BD Bioscience). Viable cells gated on the FSC/SCC were sorted on the basis of the expression of GFP, and/or staining with anti-feeder cells antibody (mEFSK4 clone) (Miltenyi).

Mouse tail fibroblasts

Mouse tail fibroblasts (dermal fibroblasts or DF) were obtained from individual, 8-10 week-old mice. Excised tails were cleaned with ethanol and placed on a falcon on ice. The tail was placed in a Petri dish and the skin was separated from the bone. The skin was cut in small pieces using scissors and forceps. The pieces were incubated in PBS with Collagenase I (2 mg/mL. Gibco) for 15 minutes at 37 °C with vortex every 5 minutes to generate a single cell suspension. The supernatant was filtered through a cell strainer (40 μm, Falcon), the digestion was repeated with the sedimented pieces and the supernatants were pooled together. For cell sorting, the pellet was resuspended in 100 μL of sorting buffer and the samples were processed as described above.

Cardiomyocytes (CM) isolation

Five days after infarct induction, CM were isolated from hearts of healthy and infarcted mice. Animals were anesthetized, their chest was open and the heart cannulated for standard Langendorff retrograde perfusion[54]. Perfusion was performed under constant pressure (60 mmHg; 37 °C, 8 min) in Ca^{2+}-free buffer containing 113 mM NaCl, 4.7 mM KCl, 1.2 mM $MgSO_4$, 5.5 mM glucose, 0.6 mM KH_2PO_4, 0.6 mM Na_2HPO_4, 12 mM $NaHCO_3$, 10 mM $KHCO_3$, 10 mM Hepes, 10 mM 2,3-butanedione monoxime, and 30 mM taurine. Digestion was initiated by adding a mixture of recombinant enzymes (0.2 mg/mL Liberase Blendzyme, Roche; 0.14 mg/mL trypsin, Invitrogen), and 12.5 μM $CaCl_2$ to the perfusion solution. When the heart became swollen (10 min), it was removed and gently divided into small pieces with fine forceps in the same enzyme solution. Heart tissue was further mechanically dissociated using 2, 1.5, and 1 mm-diameter pipettes, until all large heart tissue pieces were dispersed. The digestion buffer was neutralized using 10% FBS and 12.5 μM $CaCl_2$. CM were pelleted by gravity (20 min), the supernatant aspirated and cells resuspended in the perfusion solution containing 5% FBS and 12.5 μM $CaCl_2$. The calcium concentration was

increased by gradually adding $CaCl_2$ from 62 µM to 1 mM final concentration. Cells were collected by centrifugation and the pellet was homogenized with TRIzol™ Reagent (ThermoFisher). RNA isolation was performed according to the manufacturer's instructions.

Single-cell RNA-sequencing (scRNA-seq)

The transcriptome of isolated GFP^+-CF from 8-12 weeks old, healthy or infarcted mice (7, 14 or 30 dpi) were examined using Single Cell 3' Reagent Kits v2 (10X Genomics) according to the manufacturer's instructions. Two hearts were pooled in each scRNA-seq experiment. For experiments with *Cthrc1-KO* mice, only one heart was used. Briefly, 25,000 $GFP^+/CD31^-/CD45^-$ or $mEFSK4^+/CD31^-/CD45^-$ (for *Cthrc1-KO* mice) events were sorted in 1X PBS, 0.05% BSA and the number of cells was quantified in a Neubauer chamber. Approximately 16,000 cells were loaded at a concentration of 1,000 cells/µL on a Chromium Controller instrument (10X Genomics) to generate single-cell gel bead-in-emulsions (GEMs). In this step, each cell was encapsulated with primers containing a fixed Illumina Read 1 sequence, followed by a cell-identifying 16 bp 10X barcode, a 10 bp Unique Molecular Identifier (UMI) and a poly-dT sequence. A subsequent reverse transcription yielded full-length, barcoded cDNA. This cDNA was then released from the GEMs, PCR-amplified and purified with magnetic beads (SPRIselect, Beckman Coulter). Enzymatic Fragmentation and Size Selection was used to optimize cDNA size prior to library construction. Illumina adaptor sequences were added, and the resulting library was amplified via end repair, A-tailing, adaptor ligation and PCR. Library quality control and quantification was performed using Qubit 3.0 Fluorometer (Life Technologies) and Agilent's 4200 TapeStation System (Agilent), respectively. Sequencing was performed in a NextSeq500 (Illumina) (Read1: 26 cycles; Read2: 57 cycles; i7 index: 8 cycles) at an average depth of 50,000 reads/cell.

Bulk MARS-seq

Bulk RNA-seq was performed following a massively parallel RNA single-cell sequencing (MARS-seq) protocol adapted for bulk RNA-seq[55, 56] with minor modifications. Briefly, 3,000 to 10,000 cells were sorted in 100 µL of Lysis/Binding Buffer (Ambion), vortexed and stored at -80 °C until further processing. For cultured CF, samples were trypsinized, homogenized in 200 µL of Lysis/Binding Buffer and stored at -80 °C until further processing. Poly-A RNA was extracted with Dynabeads Oligo (dT) (Ambion) and reverse-transcribed with AffinityScript Multiple Temperature Reverse Transcriptase (RT) (Agilent) using poly-dT oligo primers (IDT) carrying a 7 bp index. Up to 8 samples with similar overall RNA content were pooled together and subjected to linear amplification via IVT using the HiScribe T7 High Yield RNA Synthesis Kit from New England Biolabs (NEB). The resulting antisense RNA was fragmented into 250-350 bps fragments with RNA Fragmentation Reagents (Ambion) and dephosphorylated with 1U FastAP (Thermo Scientific) for 15 min at 37 °C. Partial Illumina adaptor sequences[55] were ligated with T4 RNA Ligase 1 (NEB) followed by a second reverse transcription. Full Illumina adaptor sequences were added during library amplification with KAPA HiFi DNA Polymerase (Kapa Biosystems). Libraries were quantified using a Qubit 3.0 Fluorometer and their size profiles examined in Agilent's 4200 TapeStation System. Sequencing was carried out in an Illumina NextSeq 500 using paired-end, dual-index sequencing (Rd1: 68 cycles; Rd2: 15 cycles; i7: 8 cycles) at a depth of 10 million reads per sample. Between 2 and 4 biological replicates per population were performed.

Bulk mcSCRB-seq

The transcriptome of mouse CM was examined using molecular crowding single-cell RNA barcoding and sequencing (mcSCRB-seq)[57], adapted for bulk RNA-seq (this paper). A 100 ng

aliquot of total RNA was used in each experiment. Poly-A RNA was extracted with Dynabeads Oligo (dT) (Ambion) and eluted in 8 µL Tris Cl pH 7.5. Prior to denaturing at 72 °C for 3 min, 1 µL of 2 µM barcoded oligo-dT primer was added [57]. Reverse transcription was performed at 42 °C for 90 min in a mixture containing: 1X Maxima RT buffer, 7.5 % PEG 8000, 1 mM dNTPs, 5 µM of unblocked template-switching oligo[57], and 240 U of Maxima H, Minus Reverse Transcriptase (Thermo Scientific). Primer excess was removed through digestion with Exonuclease I (NEB). cDNA was purified with a 1.2X SPRI clean up (Agencourt AMPure XP, Beckman Coulter), then subjected to PCR amplification using Terra polymerase (Takara Bio) and SINGV6 primer with the following thermal cycling parameters: 3 min at 98 °C followed by 6 cycles of 15 sec at 98 °C, 30 sec at 65 °C, and 4 min at 68 °C, and a final elongation step of 10 min at 72 °C. Upon purification (1X SPRI clean-up), 0.8 ng of double stranded cDNA were tagmented using Nextera-XT Tn5 (Illumina). PCR amplification using P5NEXTPT5 primer and i7 indexed primers allowed 3' enrichment and secondary barcoding of the libraries as well as addition of Illumina adaptor sequences. Final libraries were quantified, and their profiles examined as described above. Sequencing was carried out in an Illumina NextSeq 500 using paired-end, dual-index sequencing (Rd1: 16 cycles; Rd2: 67 cycles; i7: 8 cycles) at a minimum depth of 10 million reads per sample (n=3).

Truseq RNA-sequencing

Total RNA from frozen biopsies of the different anatomical regions from swine and human hearts was isolated using TRIzol reagent (Ambion). Following mechanical homogenization with an Ultra-turrax (T10 basis Ultra-Turrax, IKA), RNA was either extracted according to the manufacturer's instructions or stored at -80 °C until processed. RNA concentration was quantified using a Qubit 3.0 Fluorometer and its quality was examined in Agilent's 4200 TapeStation System.

Swine cDNA libraries were performed using the Truseq Stranded mRNA library prep kit (Illumina). mRNA was selected using poly-dT magnetic beads followed by RT, second strand synthesis, 3' adenylation, Y shaped adaptor ligation and library enrichment. Custom Y shaped adaptors were used[58]. Final libraries were quantified, and their profiles examined as described above. Sequencing was performed in an Illumina NextSeq 500 using single-end, dual-index sequencing (Rd1: 75 cycles; i7: 8 cycles; i5: 8 cycles) at a minimum depth of 10 million reads per samples (n=19). Human RNA was subjected to rRNA depletion using Truseq Stranded Total RNA library prep Gold kit (Illumina), followed by RT, second strand synthesis, 3' adenylation, Y shaped adaptor ligation and library enrichment as described above. Final libraries were quantified, and their profiles examined as described above. Sequencing was performed in an Illumina NextSeq 500 using single-end, dual-index sequencing (Rd1: 75 cycles; i7: 8 cycles; i5: 8 cycles) at a minimum depth of 20 million reads per samples (n=21).

ATAC-sequencing

Accessible chromatin mapping was performed using FAST-ATAC-seq[59] with minor modifications. Briefly, 10,000 cells per animal (n=2 at each time point or condition) were sorted in 1X PBS, 0.05% BSA and pelleted by centrifugation at 500 rcf for 5 min at 4 °C with low

acceleration and brake settings in a pre-cooled swinging-bucket rotor centrifuge. All the supernatant but 5 µL was removed. Next, 25 µL of transposase mix (15 µL of 2x TD buffer (Illumina); 1 µL of TDE1 (Illumina), 0.25 µL of 5% digitonin (Promega); 8.75 µL of nuclease-free water) were added to the cells and the pellet was disrupted by pipetting. Transposition reactions were incubated at 37 °C for 30 min in an Eppendorf ThermoMixer with shaking at 450 rpm. Reactions were stopped at 4 °C for 5 min. In order to release tagmented DNA, samples were incubated at 40 °C for 30 min with 5 µL of clean up buffer (900 mM NaCl (Sigma), 30 mM EDTA (Millipore), 2 µL of 5% SDS (Millipore) and 2 µL of Proteinase K (NEB)). DNA was purified using a 2X SPRI beads cleanup kit (Agencourt AMPure XP, Beckman Coulter). In order to determine the total number of PCR cycles needed for library amplification, two sequential PCRs were performed using KAPA HiFi DNA Polymerase and customized Nextera PCR indexing primers (IDT)[60]. The conditions of the first PCR were: 5 min at 72 °C and 2 min at 98 °C followed by 9 cycles of 20 secs at 9 °C, 30 secs at 63 °C, and 1 min at 72 °C. Depending on the library concentration obtained in this first PCR, a second PCR (2 min at 98 °C followed by 4 to 6 cycles of 20 secs at 98 °C, 30 secs at 63 °C, and 1 min at 72 °C) was performed aiming for a library concentration in the range of 2 to 10 ng/µL. PCR products were purified using a 2X SPRI beads cleanup. Libraries were quantified and their size profiles examined as described above. Sequencing was carried out in an Illumina NextSeq 500 using paired-end, dual-index sequencing (Rd1: 38 cycles; Rd2: 38 cycles; i7: 8 cycles; i5: 8 cycles) at a depth of 80 million reads per sample.

Functional assay for TGF-β signaling pathways *in vitro*

For cell culture, CIC were resuspended in complete medium (DMEM + 10% FBS + 1% P/S + 1% L-Glu) containing 10 ng/mL of bFGF and plated in a 0.1% (w/v) gelatin-coated 6-well-plate well (one heart per well). CF were selected by attachment through washing the wells twice with PBS after overnight incubation and followed by a replacement of the complete medium plus 10 ng/mL of bFGF, lowering the concentration to 5 ng/mL (passage 1) and 2 ng/mL (passage 2) over time. After 24 hours in passage 2, cultured CF were incubated in starving conditions (complete DMEM without FBS) overnight, and then the medium was supplemented with hTGF-β1 (10 ng/mL, Peprotech), and LY294002 (40 μM, LC Laboratories) or DMSO (Sigma, used as vehicle) for 24-48 hours. To validate the effect of the inhibitor, LY294002, cell pellets were collected one hour after the treatments described above (Control (Non-treated), TGF-β, TGF-β + DMSO, TGF-β + LY294002). WT CF cultured in the four different conditions were submitted to bulk RNA-seq together with *Cthrc1-KO* CF non-treated and supplemented with TGF-β. Four biological replicates were analyzed.

In order to explore the canonical and non-canonical TGF-β signaling pathway, the effect of different inhibitory small molecules in cultured CF was examined. After 24-48 hours in passage 2, cultured CF were incubated in starving conditions overnight, and then the medium was supplemented with TGF-β (10 ng/mL), and different inhibitors or DMSO (vehicle) at the concentrations indicated in **Table II in the Supplement** for 24 hours. To validate the effect of the different inhibitors on the phosphorylation of their specific targets, cell pellets were collected one hour after the treatments described above.

In vitro Scratch Assay

GFP^+ adult CF from three mice in passage 2 were seeded in 24-well culture plates at a density of 1 x 10^4 cells/well. After 48 hours, cells were subjected to starving conditions overnight. On the following day, cells were scraped in a straight line with a pipette tip. Cells were washed twice with PBS to remove debris and serum-free medium supplemented with different treatment (control, TGF-β, TGF-β + DMSO, TGF-β + LY294002) was added to the cells. The capacity of the cells to migrate into the denuded area was tracked using time-lapse microscopy *Cell Observer Z1* (Zeiss) for 23 hours **(Movie I in the Supplement)**. Covered area was quantified using Fiji Software (1.46). Statistical significance was analyzed using a non-parametric one-way analysis of variance with a Kruskal-Wallis post-hoc test and shown as mean ± SD using GraphPad Prim 6.0 (GraphPad Software) ($p<0.05$). Between 4 and 8 biological replicates were performed per condition.

Lentiviral vectors construction

Lentiviral vector containing *Runx1* cDNA (pLenti-GIII-CMV-Runx1-GFP-2A-puro) was purchased from ABM (LV496973). Cloning vector containing *Sox9* cDNA (pcDNA3.1+/C-(K)DYK, OMu22479D) was purchased from GenSpript for posterior cloning in the lentiviral vector. Briefly, *Sox9* cDNA was amplified from the pcDNA with primers containing restriction sites for the enzymes present in the pLenti around *Runx1*, *NheI* and *XbaI,* using Platinum®Taq DNA polymerase High Fidelity (Life Technologies), followed by cloning downstream of the CMV promoter in pLenti-GII-GFP-2A-Puro after restriction digestion of the vector and the amplification fragment with *NheI* and *XbaI* to yield pLenti-CMV-Sox9-GFP. Briefly, digestion was performed for 4 hours at 37 °C and fragments were either purified with NucleoSpin Gel and PCR Clean-up column (Macherey Nagel) directly or run in an electrophoresis gel and purified from the gel band

with the same kit. Next, a ligation reaction was performed at 16 °C overnight with the digested plasmid, the insert and 0.5 µL of T4 DNA ligase (NEB). The ligation product was transformed into 25 µL of Stbl3 competent cells (Life Technologies). Clones were tested by *NheI* and *XbaI* digestion and Sanger sequencing. An empty lentiviral vector (pLenti-CMV-GFP) was constructed from the pLenti-GIII-CMV-Runx1-GFP-2A-puro by removal of the *Runx1* cDNA insert by digesting the vector with *EcoRV* (NEB) followed by purification and cloning as described above. Once cloned, constructs were confirmed by sequencing.

Lentiviral production

All lentiviruses were produced for the infection of CF by co-transfecting 5 x 10^6 HEK-293T cells with 9 µg of transcription factor (TF) of interest encoding plasmid (*Runx1, Sox9, Empty*), 6 µg psPAX2 (Addgene, #12260) and 3 µg of pMD2G (Addgene, #12259) plasmids using 40 µL of Lipofectamine2000 (Invitrogen) and 1,970 µL Opti-MEM (Gibco) in a 15 cm culture dish. Lentivirus-containing supernatant was harvested 72 hours later, filtered through a 0.45 µm HV-Durapore Stericups (SLHV033RS Millipore) and viral particles were purified and concentrated. Briefly, the filtrates were transferred to Ultra-Clear Beckman tubes (Beckman Coulters) and centrifuged at 26,000 rpm for 2.5 hours at 4 °C in an Ultracentrifuge (Beckman optima LE-80K). The pellet was resuspended in an appropriate volume of PBS depending on the number of culture dishes. The lentivirus stocks were snap-frozen in 100 µL aliquots and stored at -80 °C. Vector titers were determined by serial dilutions of the concentrated lentivirus on HEK-293T cells in cultured medium supplemented with 8 µg/mL polybrene (Sigma) followed by flow cytometry analysis 72 hours after infection.

Lentiviral transduction

For transduction, cultured CF were infected at passage 1 with virus stock in culture medium supplemented with 8 µg/mL polybrene (Sigma). After addition of the virus, cell plates were centrifuged at 2,000 rpm for 90 min at 32 °C without acceleration or break in a pre-warmed swinging-bucket rotor centrifuge. After 24 hours the infection was repeated. The cells were incubated for a total of 72 hours from the first infection and then subcultured from passage 1 to 2. After four days in culture (passage 2) cells were trypsinized and sorted in 100 µL of Lysis/Binding Buffer, vortexed and stored at -80 °C until further processing for RNA-seq. Cells infected with the empty vector was used as control. Between 4 and 6 biological replicates were analyzed.

Immunofluorescence

Mice adult hearts were excised and washed in PBS, fixed in 4% fresh paraformaldehyde (Sigma). Some specimens were then cryoprotected in sucrose (Sigma), embedded in OCT compound (VWR), frozen in dry ice and stored until sectioning. For RCF specific marker staining additional specimens were dehydrated, paraffin embedded, sectioned and rehydrated before the staining. Pig's biopsies of infarct, border and remote zones were paraffin embedded, sectioned and rehydrated. Five-to-ten micron thick sections were washed in Tris-PBS (TPBS), non-specific IgG binding sites were blocked with 8% goat serum (Dako), 1% BSA, and 0.1% Triton X-100 (Sigma) followed by incubation overnight with the corresponding primary antibodies at 4 °C **(Table III in the Supplement)**. CTHRC1 and anti-GFP staining required an antigen retrieval step using 100 mM citrate buffer pH6.3 (Sigma) before blocking. The sections were then washed and incubated with the appropriate fluorescence-conjugated secondary antibodies for 1 hour before mounting **(Table IV in the Supplement)**. Negative controls were included by omitting the primary antibody. Cell

nuclei were counterstained with 4',6-diamidino-2-phenylindole (DAPI, Sigma) or Hoechst 33258 (Invitrogen, 1 µg/mL for 5 min). All images were captured in a Zeiss LSM 510 800 (Zeiss) or Leica SP8 confocal laser microcopy (Leica). Sections from the angiotensin II model and human tissues were stained for CTHRC1 (clone Vli55, 100 ng/mL) followed by visualizing with diaminobenzidine (DAB) as described previously[61]

Morphometric assay: quantification of GFP^+ and $CD200^+/GFP^+$ cells

The quantification of the area occupied by GFP^+ or $CD200^+/GFP^+$ was performed in stained sections from healthy myocardium and infarct tissue at 7dpi in three different mice. For infarcted hearts, between 1 and 5 pictures from infarct, border and remote zones (IZ, BZ, RZ, respectively) were taken from two different transverse sections per animal, located 150-200 µm from each other. For healthy hearts, the free wall of the left ventricle (LV) and right ventricle (RV) were selected. The number of pixels from GFP^+ (green) or $CD200^+/GFP^+$ (yellow) areas and $DAPI^+$ nuclei (blue) were determined using user manual macros in Fiji software (1.46). For graphic representation, the percentage of the area of GFP^+ or $CD200^+/GFP^+$ cells were calculated for each region. The results were analyzed using a non-parametric one-way analysis of variance with a Kruskal-Wallis post-hoc test using GraphPad Prism 6.0 (GraphPad Software) ($p<0.05$).

Kaplan-Meier survival curve

To analyze animal survival, MI was induced in groups of *Cthrc1-KO* and *Cthrc1-WT* animals (n=10 per group). Survival of animals was carefully monitored once a day for seven days after MI. Log-rank Mantel-Cox test was used to determine statistical difference between the survival curves of the two groups of animals.

Quantification of Collagen deposition

Picrosirius red (PSR)-polarization detection of collagen fibers in tissue sections was used to determine collagen deposition in left ventricular walls at 3dpi *Cthrc1-KO* and *Cthrc1-WT* mice hearts (n=5 per group). Slides were incubated with Sirius Red (0.1%) dissolved in aqueous saturated picric acid overnight. PSR-stained sections were imaged under polarized light using an Axioskop 40 microscope (Zeiss). Deposition of collagen fibers were analyzed using Fiji software (1.46). Statistical significance was analyzed by an unpaired *t*-test using Graphpad Prism 6.0 (GraphPad Software).

Western Blot

Passage 2 CF were trypsinized, spun at 1,200 rpm for 5 min, pellets frozen in liquid nitrogen and stored at -80 °C until further processing. Total protein was extracted using 1% Triton-X100 in a cocktail of protease inhibitors (EDTA, sodium orthovanadate, sodium fluoride, Sigma), and quantified using a BCA Protein Assay Kit (ThermoFisher Scientific). Proteins were separated by electrophoresis on 10-12% acrylamide gel and blotted onto nitrocellulose membranes (Bio-Rad, 0.45 µm). Membranes were blocked using the corresponding blocking solution (5% non-fat dry milk powder or 5% BSA, 0.05% Tween-20, 1 mM Tris-HCl, 15 mM NaCl) for 1-3 hour at room temperature **(Table V in the Supplement)**. Membranes were washed using TBS-0.05% Tween-20 solution. Lumigen ECL Ultra (TMA-6, Beckman Coulter) was used for the development of the signal (5 min at room temperature) after incubation with the corresponding HRP-conjugated secondary antibody **(Table V in the Supplement)**. *ChemiDoc™ Imaging Systems* (Bio-Rad) was used for image acquisition. Three separate replicates were analyzed.

RNA isolation and qPCR

For cultured mice CF, samples were collected in TRIzol (Ambion) reagent after 24 hours of treatment, stored at -80 °C until processing, and isolated according to manufacturer's instructions. Total RNA from frozen biopsies of the different anatomical regions of 21 infarcted pigs at 180dpi were isolated as described above, using TRIzol reagent. RNA purity (A260/A280 ratio ≥ 1.8) and concentration were measured using the Nanodrop (Thermo Fisher). Total RNA was stored at -80 °C. Prior to cDNA preparation, genomic DNA was removed using the Thermo Scientific Dnase treatment kit (Thermo Fisher). For mouse CF samples, a total of 500 ng of RNA was converted into cDNA as per manufacturer's instructions (Takara). For swine samples, a total of 1 µg of RNA

was converted into cDNA using 125 µM of anchored oligo-dT primers (IDT) and Superscript II reverse transcriptase (RT, Life Technologies).

For each qPCR, the cDNA equivalent to 5 ng RNA was used. The qPCR reactions contained power SYBR green PCR master mix (Applied Biosystems) and an equimolar primer mix (0.8 µM). The amplification protocol consisted of 2 min at 50 °C, 10 min at 95 °C, followed by 40 cycles of 15 secs at 95 °C, and 1 min at 60 °C, and completed with a standard melting curve protocol (15 secs at 95 °C, 1 min at 60 °C and 15 secs at 95 °C). The melting curve analysis (ViiA TM 7 Real-Time PCR system, Applied Biosystems) and size fractionation by agarose gel electrophoresis were used to confirm amplification of the expected products[62]. The expression level of mRNA (N_0) was obtained from the extracted raw data using *LinRegPCR* program[63]. Systematic differences induced by RT reactions among samples as well as systematic differences between qPCR runs were removed using *Factor-qPCR* program[64]. *Gapdh* and *Rpl4* for mouse samples, and *HPRT1* and *RPL4* for porcine samples were selected as reference genes after a stability analysis as described previously[65, 66]. Primer sequences were designed using *primer3*, *BLAST* (NIH) and *oligo analyzer* (IDT) software and are summarized in **Table VI in the Supplement**. Graphs were generated and statistical analysis was performed by a two-ways ANOVA in *GraphPad* Prim version 6.0 (GraphPad Software) ($P<0.05$). N_0 values for periostin (*POSTN*), collagen 1α1 (*COL1α1*) and collagen triple helix repeat containing 1 (*CTHRC1*) quantified in the infarcted area in pigs were correlated with the corresponding value of the ejection fraction (EF) (%). R = Spearman correlation coefficient.

scRNA-seq analysis

Sequenced libraries were demultiplexed, aligned to the mouse transcriptome (*mm10*) and quantified using *Cell Ranger* (2.0.0) from 10X Genomics. The output of the pre-processing pipeline consisted of gene expression matrices per cell. Further computational analysis was performed using *Seurat* (2.3.4). Cells were subjected to quality control filters based on the number of detected genes, number of UMIs and proportion of UMIs mapped to mitochondrial genes per cell. The thresholds for each of the single cell libraries were selected based on the distribution of the previously mentioned variables and visual inspection of quality control scatter plots **(Figure III in the Supplement)**. Using these parameters, 7,079 (healthy myocardium), 10,448 (7dpi), 8,337 (14dpi), 6,805 (30dpi) and 4,189 (*Cthrc1-KO*) cells were retained.

Each single cell dataset was subjected to normalization, identification of highly variable genes and removal of unwanted sources of variation as described previously[67]. To integrate each of the analyzed time points, an analysis based on canonical correlation analysis (CCA)[68] was used. Briefly, genes identified among the top 1,000 most variable genes in at least 3 datasets were selected as input for CCA, and 30 canonical vectors were calculated. Sixteen CCA components were selected by visual inspection of biweight correlation plots. Cells whose variance explained by CCA was < 2-fold compared with principal component analysis (PCA), were removed. The integrated single cell dataset was subjected to unsupervised clustering, with a resolution parameter of 0.8. Non-linear dimensional reduction was performed using t-distributed stochastic neighbor embedding (t-SNE)[69]. All differential expression tests were done using the function *FindAllMarkers* with default settings. Up-regulated genes detected in a fraction of 0.2 in either of the populations were deemed significant (p value < 0.01).

Bulk RNA-seq analysis

Samples were demultiplexed using Illumina *bcl2fastq* software (1.2.4) and aligned with the mouse (*mm10*), swine (*Sscrofa 11.1*) and Human (*GRCh38*) genome with STAR (2.6.1)[70] setting the parameters to default values. Quantification and generation of gene expression matrices were performed with the function *featureCounts,* implemented in the R [71] package *Rsubread*[72]. The *ensembl* transcriptomes (*GRCm38.91, Sscrofa11.1.93, and GRCh38.92*) were used as reference for gene annotation. Before statistical analysis, the function *filterbyExpr*, implemented in the R package *edgeR*[73], was used to determine genes with enough counts for further analyses. Data transformation, normalization, and testing for differential expression was performed with DESeq2[74].

For lentiviral experiments, samples for vehicle were included as covariates and the test was performed between samples with treatment and controls. For RNA-seq analyses from the set of cultured CF using TGF-β, the effect of the vehicle was also included in the analysis design.

Pathway and gene ontology analysis

To perform enrichment analysis for signaling pathways and gene ontology categories, the R package *clusterProfiler*[75] was used. For pathway enrichment, we used the Reactome[76], and for gene ontology categories, we centered on the biological process ontology. Hypergeometric tests were performed and p-values were corrected using the Benjamini-Hochberg method[77], setting the p-value cutoff at 0.01 and the q-value cutoff at 0.05. The remainder of the parameters were set to default values. Furthermore, we used *enrichR* as an interface to the *Enrichr*[78] database to perform additional tests for enrichment. In this case, the categories *GO_Biological_Process_2018 and*

BioPlanet_2019[79] were selected. Results were sorted according to the *enrichr* combined score, implemented in the *EnrichR* R package.

RNA Velocity and latent time analysis

Single cell data sets were analyzed using *velocyto*[25] (0.17.17) and one loom file was generated for each condition (healthy, 7dpi, 14dpi and 30dpi) using the BAM files generated with *CellRanger* (2.0.0) from 10X Genomics. Further processing to obtain trajectories was performed using *scVelo*[26] (version 0.1.26) implemented in *Python* (3.7.2). Each dataset was processed in *Seurat* (2.3.4) and the results were imported to *Python* as loom files. For trajectories, the data was filtered and normalized using default parameters. The dynamical model was used to estimate RNA velocities. Latent time values per condition and per cell were generated with the function *recover_latent_time* implemented in *scVelo*. The stream plots were generated in the t-SNE reductions retrieved with *Seurat* R package.

Integration of *Col1α1-GFP* mouse model and Farbehi *et al.* (2019)

Single cell dataset from Farbehi *et al.* [12] was downloaded from ArrayExpress with accession number E-MTAB-7376 and processed according to the described parameters using *Seurat* (3.0). The *Col1α1-GFP* dataset was processed as described previously. Integration of both datasets was performed using the functions *FindIntegrationAnchors* and *IntegrateData* implemented in *Seurat* (3.0)[80]. Dimensional reduction was performed and 60 PCA components were retrieved. Only 25 where kept for unsupervised clustering and UMAP visualization **(Figure V in the Supplement)**.

Transcription factor enrichment analysis

To identify TF with target genes enriched among RCF markers, we performed enrichment using the *Enrichr*[78] software and gene sets *TRANSFAC*_and_*JASPAR_PWMs ENCODE_TF_ChIP-seq_2015 ENCODE*_and_*ChEA_Consensus_TFs*_from_*ChIP*-X *ChEA_2016*. P-values were adjusted using the R function *p.adjust* with method "BH". RCF markers with adjusted p-value < 0.05 were used in the enrichment.

Network betweenness analysis

To identify upstream regulators of gene expression changes mediated by TGF-β receptors, we generated a network spanning protein interactions and transcription factor-gene interactions. We downloaded transcription factors-genes interactions from *ENCODE*, *CHEA*, *ESCAPE*, *MotifMap*, *Transfac* using the *Harmonizome*[81] database (July 3, 2017). In addition, we downloaded the TTRUST[82] database (July 3, 2017). Gene names were mapped using NCBI (ftp://ftp.ncbi.nlm.nih.gov/gene/DATA/gene_info.gz (downloaded on October 10, 2016). Then the TF with the strongest evidence were connected to each gene. A total of 745 highly biased genes were connected to more than 25 TF and were thus removed from the analysis.

Protein interactions were obtained from the *SIGNOR*[83] database (downloaded July 3, 2017). We deleted interactions annotated as "transcriptional regulation", "guanine nucleotide exchange factor", "transcriptional activation", "post transcriptional regulation", or "transcriptional repression". Then, we deleted nodes that do not correspond to proteins, their complexes or families by retaining only nodes with *UniProt*[84] or *SIGNOR* ID.

Betweenness analysis was performed in R (3.4.0) using the *igraph* package (1.1.2). All shortest paths from TGF-β receptors to a given gene list were obtained using the *igraph* function

all_shortest_paths. Then, the number of occurrences of each protein in these paths was calculated and used as a measure of graph betweenness. Finally, betweenness of cluster B was compared to betweenness of markers of other clusters.

ATAC-seq analysis

ATAC-seq libraries were demultiplexed using Illumina *bcl2fastq* software. The generated fastq files were aligned to the mouse (*mm10*) genome using bowtie2 (2.3.1)[85] with the parameters *--no-discordant, --no-mixed, --very-sensitive* and *-X 1000*. The generated SAM files were sorted and converted to BAM files using *samtools* (1.6). Peak calling was done using MACS2 (2.1.0)[86] with the parameters *-q 0.5, -B, --call-summit* and *--keep-dup=all*. Detected peaks were annotated to the *UCSC mm10* known gene database (3.4.4) using *ChIPseeker*[87] R package. Peak quantification was done using *featureCounts* implemented in Rsubread[72]. Data transformation, normalization and testing for differential open chromatin regions (OCR) was performed with DESeq2[74]. We extracted differentially accessible regions between $CD200^+$ and $CD200^-$ by selecting the ones with Adjusted P.value < $1e^{-3}$ and LogFC > 0. Only distal regions (|Distance to TSS| > 3kb) were kept for further analyses. The selected regions were subjected to Motif Enrichment Analysis using *HOMER*[88] (4.9). Novel and known motifs were identified using the function *findMotifsGenome.pl* with the parameter-*size 200*. All other parameters were set to default values.

Identification of candidate Transcription Factor global regulators

Peak clustering was conducted on normalized and z-scaled data using *mclust*[89] library in R[71]. As a result, three clusters were identified: Cluster 1 (accessible peaks only in $CD200^-$ cells), Cluster 3 (accessible peaks only in $CD200^+$ cells) and Cluster 2 (undefined). Cluster 2 was eliminated in the downstream analysis. Motif identification was conducted using *HOMER*[88] with the *mm10*

reference genome (*findmotifsGenome*). Once enriched motifs were identified and annotated, peaks were scanned (*scanMotifGenomeWide*). After using *HOMER* information to map motifs to TF, a total of 277 TF were identified. For each TF, an enrichment analysis was conducted using a Fisher test; we investigated if regions associated with a given TF were over-represented in one of the identified clusters: Cluster 1 (enriched in $CD200^-$) *versus* Cluster 3 (enriched in $CD200^+$). An adjusted p-value < 0.05 was defined as significant. Additionally, for each TF we computed the percentage of open chromatin regions (OCR) associated.

Transcription factor derived Gene-Set analysis

For each TF, a gene set analysis (GSA) was performed in order to identify biological differences in the TF regulation between $CD200^+$ and $CD200^-$ cell types. For each TF selected (n=16) we conducted the following steps:

1) Select the open chromatin regions related with the TF and mapped to promoter areas (areas of DNA that leads to initiation of transcription of a gene)
2) Associate each OCR to a $CD200^+$ or $CD200^-$ depending on the clustering enrichment; the regions enriched in cluster 1 were classified as $CD200^+$ and the regions enriched in cluster 3 as $CD200^-$.
3) Generate a gene-cell type associated based on the annotation of OCR ($CD200^+$ and $CD200^-$).
4) Run the GSA in each case.

An adjusted p-value < 0.01 was considered as significant. Finally, the significant GO associated to each cell type for each TF were compared (**File I in the Supplement).**

Data resources

All single-cell RNA-seq, bulk RNA-seq, and ATAC-seq data are available in a SuperSeries at NBCI's Sequence Read Archive database under accession number GSE132146.

SUPPLEMENTAL TABLES

Supplementary Table I: List of antibodies used in FACS and cell sorting.

Name	Clone	Supplier	Dilution
CD11b-APC/Cy7	M1/70	BioLegend, 101225	1:100
CD31 (PECAM)-APC	MEC 13.3	BD Pharmingen, 551262	1:200
CD31 (PECAM)-PE	MEC 13.3	BD Pharmingen, 555027	1:200
CD31-APC	390	BioLegend, 102409	1:200
CD45-BV510	30-F11	BioLegend, 103137	1:200
CD45-PE	30-F11	eB 12-0451-81	1:200
CD45-PerCP	30-F11	BD Pharmingen, 557235	1:200
CD45-FITC	30-F11	BioLegend, 103107	1:100
CD90.2-PE/Cy7	53-2.1	eB 25-0902-82	1:100
CD140a-PE/Cy7	APA5	eB25-1401-80	1:100
CD146-PE/Cy7	ME-9F1	Biolegend 134714	1:100
CD200-BV421	OX-90	BD Bioscience 565547	1:100
Anti-feeder cells-APC	mEFSK4	Miltenyi, 130-102-302	1:50
Anti-CTHRC1 serum	Vli08G09	Maine Medical Center Research Institute	1:100
Anti-rabbit IgG- PE	-	Jackson Inmunoresearch, 111-116-144	-

Supplementary Table II: List of small molecules related to both canonical and non-canonical TGF-β signaling pathway used in the *in vitro* experiments.

Pathway	Small molecule	Target	Concentration	Supplier
TGF-β1 signaling (general)	SB-431542	TGF-β R1 (ALK5)	1 - 10 μM	Calbiochem, 616461
Canonical TGF-β	SIS3	Smad3	0.5 – 5 μM	Calbiochem, 566405
Non-canonical TGF-β	LY294002	PI3K/Akt	10 – 40 μM	Calbiochem, 440202
	PD-98059	ERK	10 – 40 μM	Calbiochem, 513000
	Akti	Akt	0.5 – 50 μM	Calbiochem, 124018
	SB-205380	p38	5 – 20 μM	Calbiochem, 559389
No TGF-β signaling	Itraconazole (ITA)	-	0.25 – 1 μM	Calbiochem, 419825

Supplementary Table III: List of primary antibodies used in immunohistochemistry.

Name	Clone	Supplier	Dilution
CD31 (PECAM)	MEC 13.3	BD Pharmingen, 550274	1:100
CD45-PE	30-F11	eB 12-0451-81	1:100
CD200-BV421	OX-90	BD Bioscience, 565547	1:200
Collagen-I	pAb	Rockland, 600-401-1035	1:200
CTHRC1	Vli55	Maine Medical Ctr. Res. Institute	100 ng/ml
FMOD	pAb	Invitrogen, PA526250	1:100
DDAH1	pAb	Invitrogen, PA552278	1:50
GFP	pAb	Abcam, ab13970	1:100
Sarcomeric α-Actin (αSA)	EA-53	SIGMA, A7811	1:100
Cardiac Troponin 1 (TnT)	4C2	GeneTex, GTX10231	1 ug/ml
Cardiac Troponin T (cTN)	13-11	Invitrogen, MA5-12960	1:100
Periostin (POSTN)	pAb	Abcam, ab79946	1:100
Actin, α-Smoth muscle (αSMA)	1A4	Sigma, A5228	1:200

Supplementary Table IV: List of secondary antibodies used in immunohistochemistry.

Name	Supplier	Dilution
Anti-Chicken 488	Jackson Inmunoresearch, 703-545-155	1:200
Anti-Rabbit 488	Invitrogen, A11008	1:100
Anti-Rabbit 594	Invitrogen, A21207	1:200
Anti-Rat 594	Invitrogen, A11007	1:200
Anti-Mouse 488	Invitrogen, A21202	1:200
Anti-Mouse 546	Invitrogen, A11003	1:100
Anti-Mouse 647	Invitrogen, A31571	1:200
Anti-Rabbit 647	Invitrogen, A31573	1:200
Anti-Rat 647	Invitrogen, A21247	1:200

Supplementary Table V: List of primary and secondary antibodies used in Western Blot.

Primary antibodies	Clone	Supplier	Dilution
Phospho-Akt (Ser473)	pAb	Cell Signaling, 9271	1:1000
Total Akt	pAb	Cell Signaling, 9272	1:1000
Phospho-p44/42 MAPK (Erk1/2) (Thr202/Tyr204)	pAb	Cell Signaling, 9101	1:1000
Total p44/42 MAPK (Erk1/2)	pAb	Cell Signaling, 9102	1:1000
Phospho-p38 MAPK (Thr180/Tyr182)	D3F9	Cell Signaling, 4511	1:1000
Total p38 MAPK	pAb	Cell Signaling, 9218	1:1000
Phospho-Smad3 (S423 + S425)	EP823Y	Abcam, ab52903	1:1000
Phospho-Smad2	D8591	Maine Medical Ctr Res. Inst.	1:3000
Total Smad2/3	pAb	Cell Signaling, 3102	1:1000
Secondary antibodies	**Clone**	**Supplier**	**Dilution**
Donkey HRP-conjugated anti-rabbit	pAb	GE Healthcare, NA934	1:10000

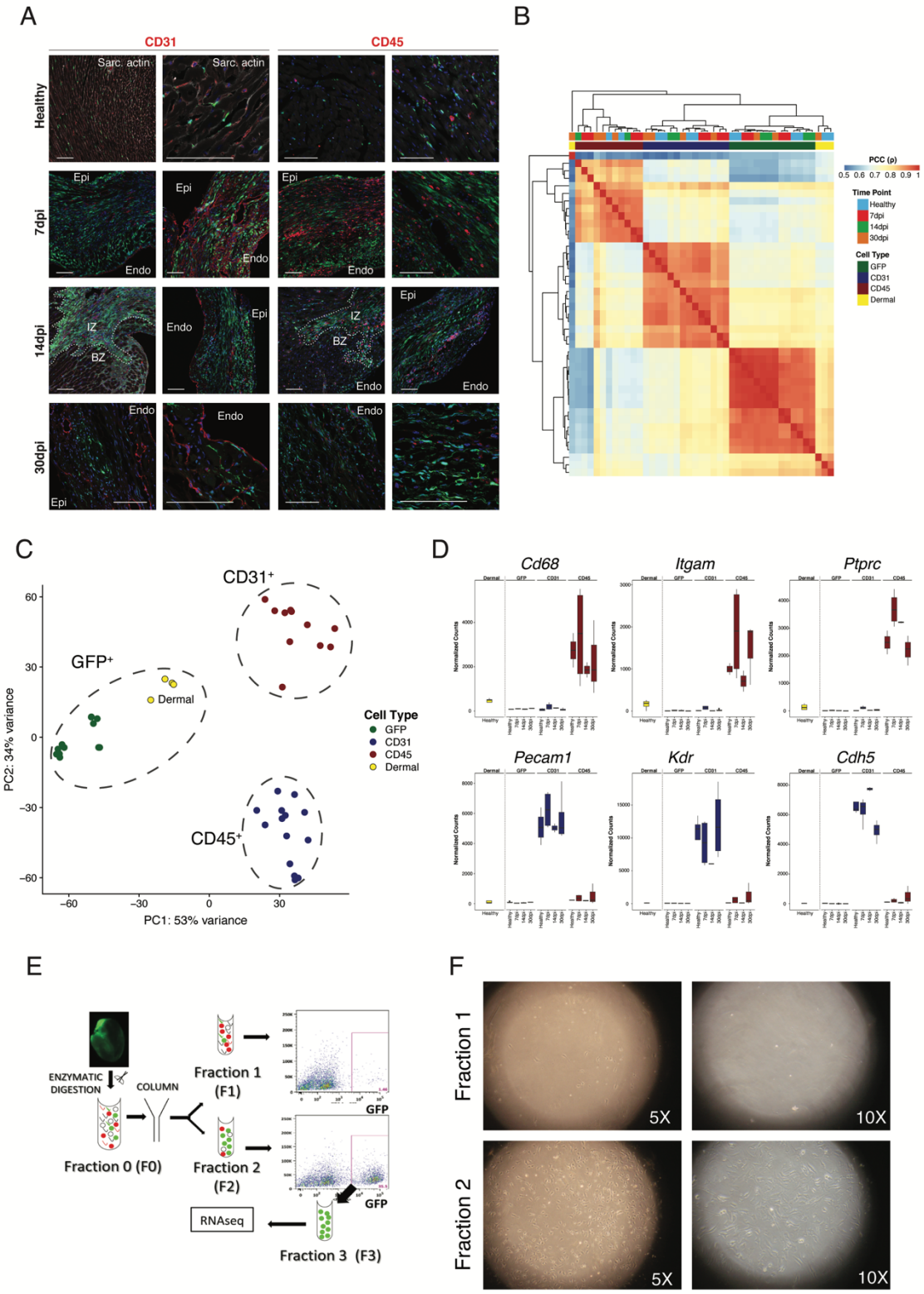
Supplementary Table VI: List of primers used in murine and swine qPCR analysis

Name	Fwd	Rev
Gapdh	CTCCCACTCTTCCACCTTCG	GCCTCTCTTGCTCAGTGTCC
Rpl4	GCCGCTGGTGGTTGAAGATAA	CGTCGGTTTCTCATTTTGCCC
Cthrc1	GCTGTCAGCGCTGGTATTTT	ACCCAGATGGCCACATCTAC
Ddah1	CACCGAACCCCAGAAGAGTA	CCGTCCACCTTTTCCATCT
Col1α1	AGGCGAAGGCAACAGTCG	TTTACACGAAGCAGGCAGGG
Col3α1	GCCTCCCAGAACATTACATACC	CTTGCTCCATTCCCCAGTGT
Lox	CCCGACCCCTACTACATCCA	AGTCTCTGACATCCGCCCT
Postn	TTCGTGGCAGCACCTTCAAA	GTCACCGTTTCGCCTTCTTT
Acta2	ATGGAGTCAGCGGGCATC	CGTTCTGGAGGGGCAATGAT
HPRT1	CCCAGCGTCGTGATTAGTGAT	TCCAGCAGGTCAGCAAAGAA
RPL4	GTTGGCATCGCAGAGTGAAC	CTCATTCGCTGAGAGGCA
COL1α1	CCTCAAGAGAAGGCTCACGA	TTGGTTGGGGTCAATCCAGT
CTHRC1	GGTCGGGATGGATTCAAA	TTAGGGCACTGTTGGAACGC
POSTN	TGTGCCAACCAATGATGCCT	ACAGGTATGTCTGCTGGGTAG

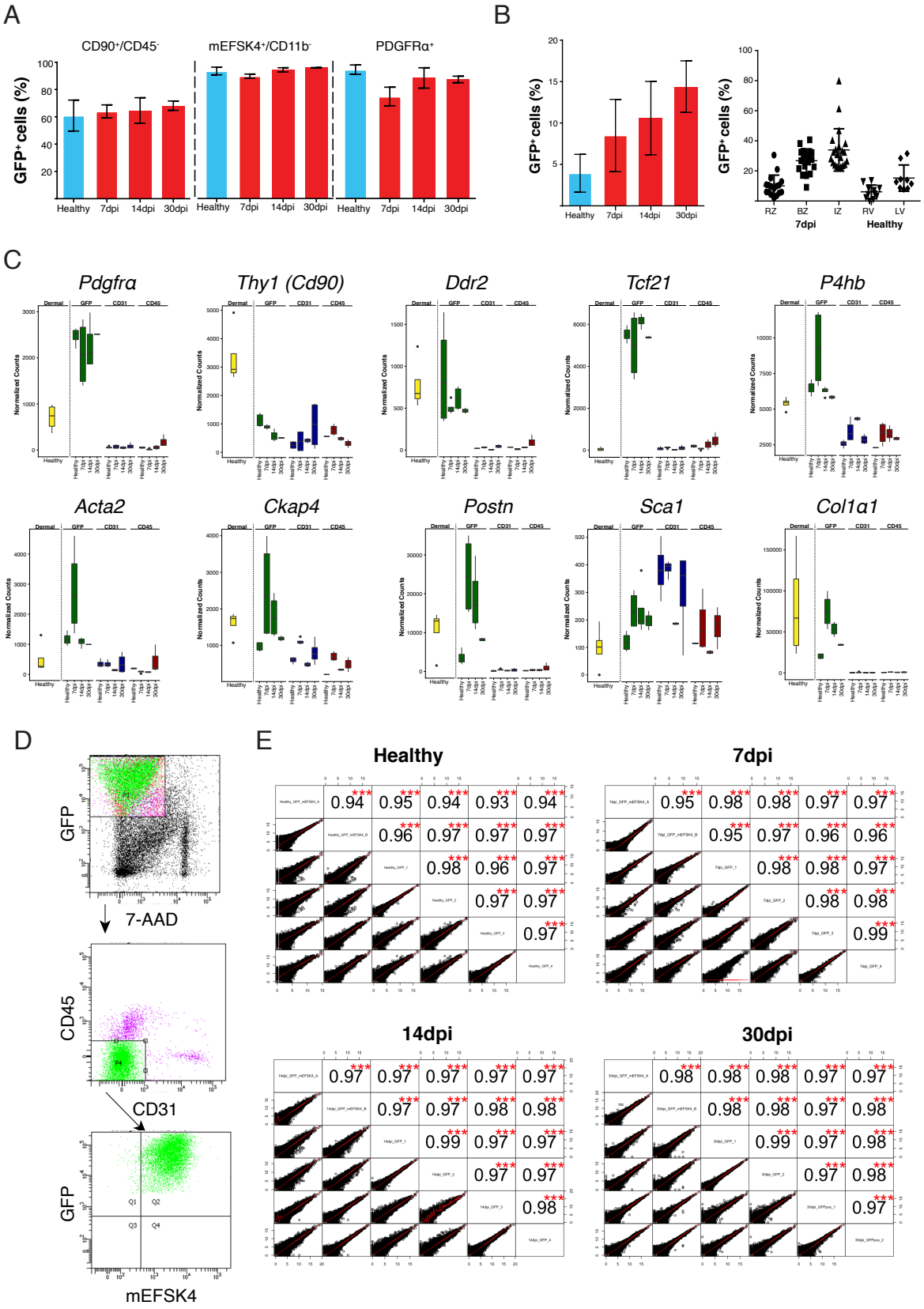
Supplementary Movie I: *In vitro* Scratch Assay
Representative movies of wound healing assay during 23 h of culture of GFP^+ adult CF in control conditions (no treatment) **(A)**, treated with TGF-β **(B)**, with TGF-β + DMSO **(C)**, and TGF-β + LY294002 **(D)**.

Supplementary File I: GSA_results.xlsx. Results of GSA of each TF analyzed

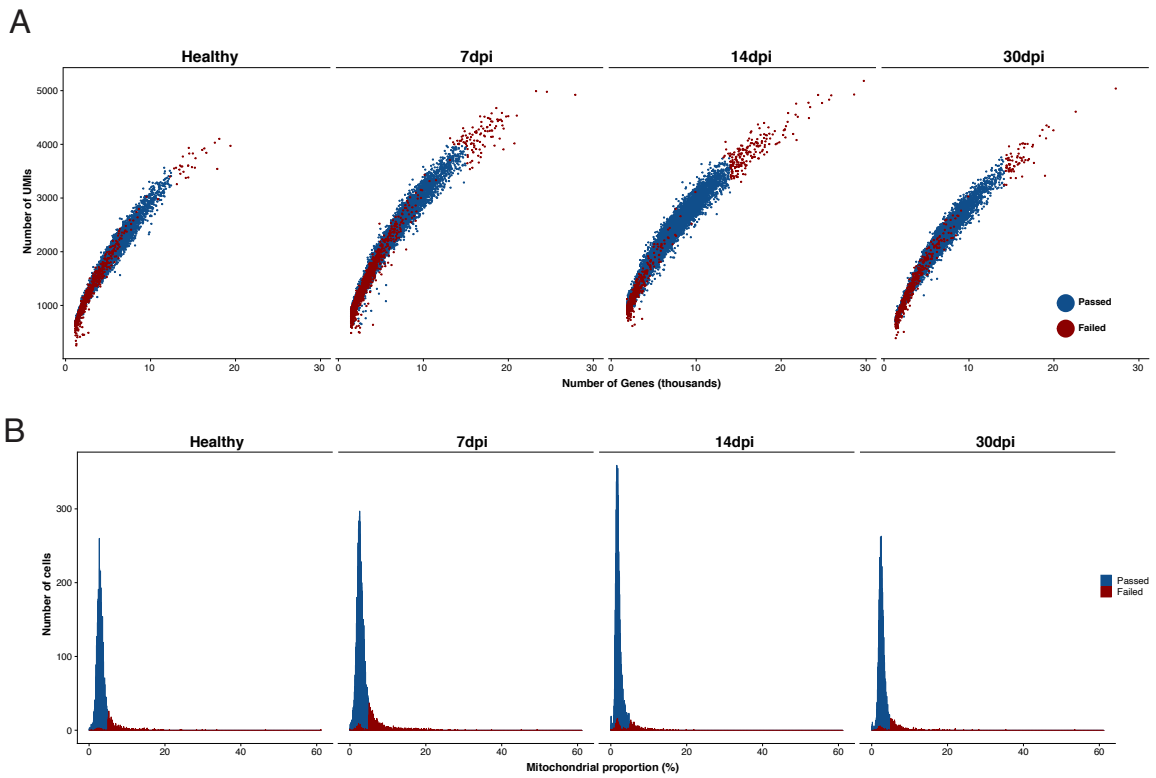
SUPPLEMENTAL FIGURES AND FIGURE LEGENDS



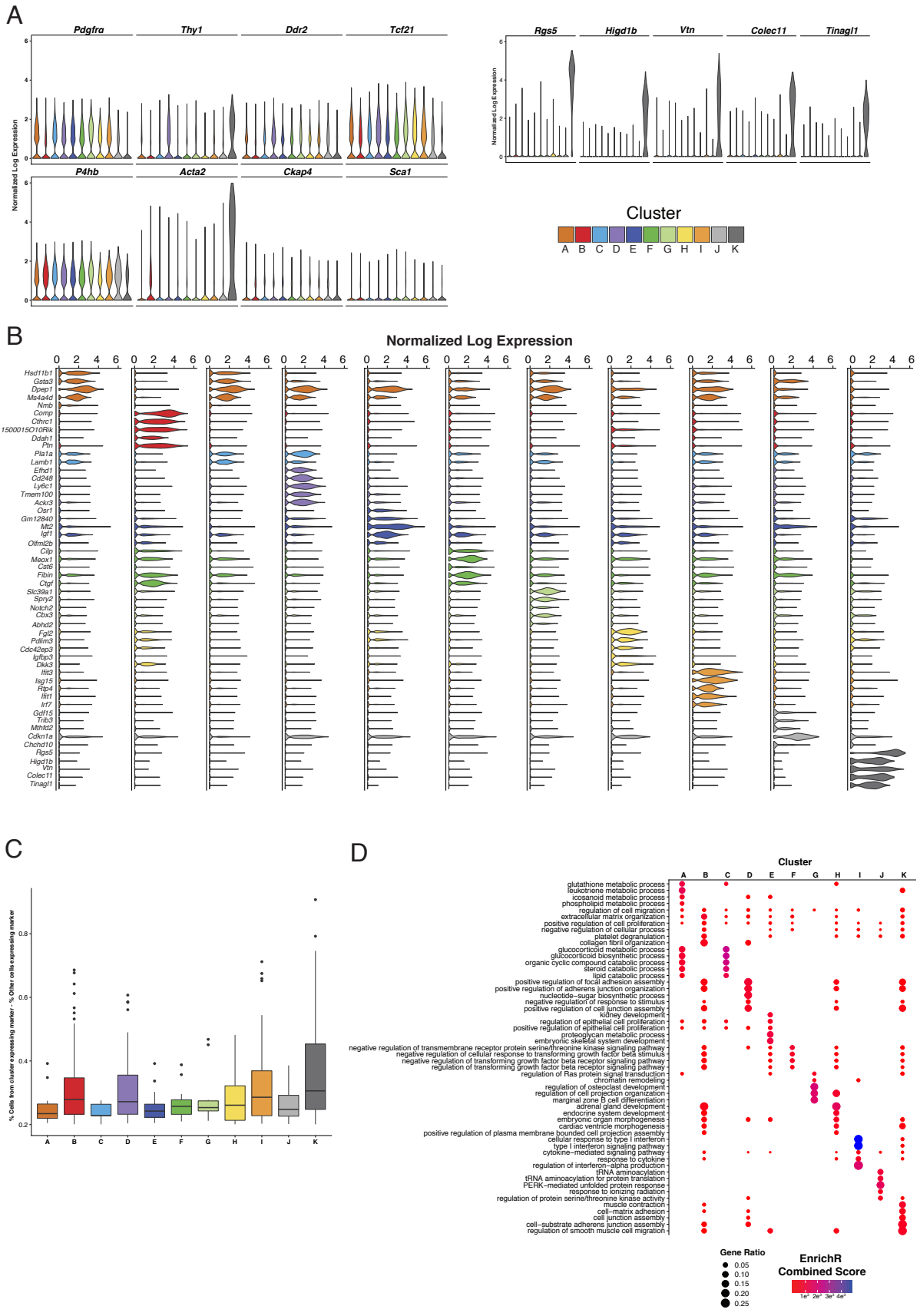
Supplementary Figure I: Characterization of cardiac interstitial cells from *Col1α1-GFP*. (A) Immunofluorescence analysis of GFP (green), CD31 or CD45 (red), α-sarcomeric actin (grey) and DAPI/nuclei (blue) in the LV in healthy myocardium, and at 7, 14 and 30 dpi. Dotted lines delimit IZ from BZ. **(B)** Correlation heatmap of transcriptomic profiles between different cell fractions isolated from healthy myocardium and along infarction: $GFP^{+}/CD31^{-}/CD45^{-}$ (green), $GFP^{-}/CD31^{+}/CD45^{-}$ (blue), $GFP^{-}/CD31^{-}/CD45^{+}$ (red) CIC, and $GFP^{+}/CD31^{-}/CD45^{-}$ (yellow) from the tail of healthy mice. **(C)** PCA scatter plot of CIC (CF, endothelial, bone marrow-derived cells) and dermal samples subjected to transcriptomic profiling. **(D)** Boxplot of expression levels (mean ± SD) of traditional markers for bone marrow-derived cells (*Cd68, Itgam, Ptprc)* and for endothelial cells (*Pecam1, Kdr, Cdh5)* in different cell fractions isolated from healthy myocardium and 7, 14 and 30 dpi. **(E)** Diagram showing the technical procedure performed to obtain a single-cell population containing almost the complete set of $GFP^{+}/CD31^{-}/CD45^{-}$ cells. **(F)** Pictures showing culture of different fractions of cell isolation corresponding to the diagram in (E) after 4 days. Fraction 1: cells not retained by column (negative fraction). Fraction 2: retained portion of cells (positive fraction). CIC = cardiac interstitial cells; Endo = endocardium; Epi = epicardium; BZ = border zone; IZ = infarct zone; LV = left ventricle; RV = right ventricle. Scale bars: 100 μm.


A
CD90+/CD45-
mEFSK4+/CD11b-
PDGFRα+
GFP+ cells (%)
Healthy
7dpi
14dpi
30dpi
B
GFP+ cells (%)
RZ
BZ
IZ
RV
LV
7dpi
Healthy
C
Pdgfra
Thy1 (Cd90)
Ddr2
Tcf21
P4hb
Acta2
Ckap4
Postn
Sca1
Col1a1
Normalized Counts
Dermal
GFP
CD31
CD45
D
GFP
7-AAD
CD45
CD31
mEFSK4
E
Healthy
7dpi
14dpi
30dpi

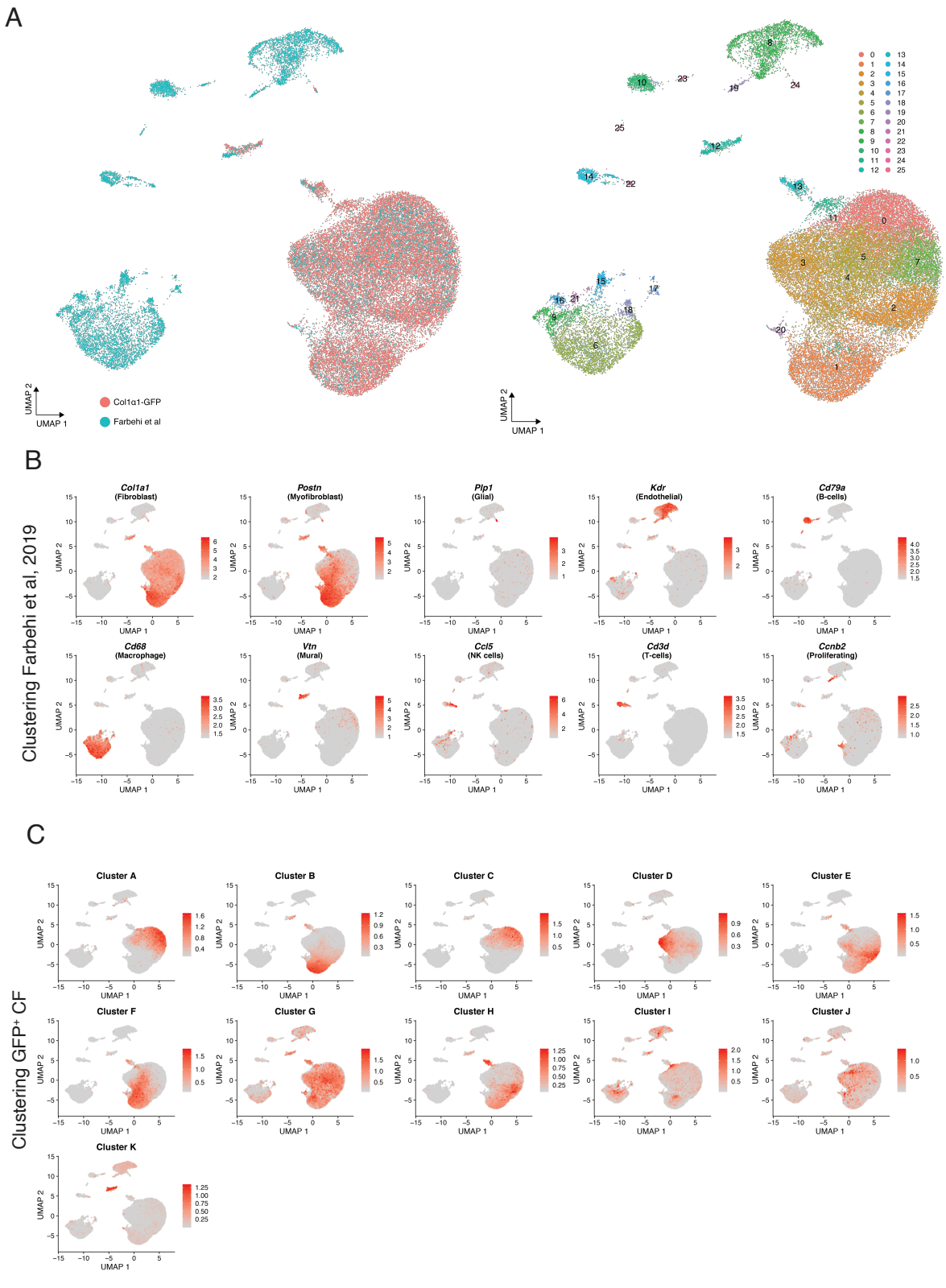
Supplementary Figure II: *Col1α1*-GFP^{+}/CD31^{-}/CD45^{-} cells are true cardiac fibroblasts. (A) Barplots of the percentage of GFP^{+}/CD31^{-}/CD45^{-} cells that express classical cell surface markers for CF (CD90 (*Thy1*), mEFSK4, PDGFRα (CD140a)). **(B)** Barplots of the percentage of GFP^{+}/CD31^{-}/CD45^{-} cells in healthy myocardium and at 7, 14 and 30 dpi (left). Quantification of GFP^{+} cells at 7dpi and in healthy myocardium (right). **(C)** Boxplotf of the expression levels (mean ± SD) of traditional markers for non-activated (*Pdgfrα, Thy1, Ddr2, Tcf21, P4hb, Col1α1*) and activated CF (*Acta2, Postn, Ckap4, Sca1*) in different cell fractions. **(D)** Representative gating for isolation of mEFSK4^{+}/GFP^{+}/CD31^{-}/CD45^{-} cardiac interstitial cells. *From above to below*: gating strategy for GFP^{+} viable cells (GFP^{+}/7AAD^{-}), for CD45 and CD31 vs GFP^{+} cells, and for GFP^{+}/mEFSK4^{+} cells **(E)** Pearson correlation of transcriptomic profiles between GFP^{+} and mEFSK4^{+} in healthy and infarcted mice (7, 14 and 30 dpi). * $p \leq 0.05$, ** $p \leq 0.01$, *** $p \leq 0.001$. RZ = remote zone; rest of abbreviations as in Supplementary Figure I.


A
Healthy
7dpi
14dpi
30dpi
Number of UMIs
Number of Genes (thousands)
Passed
Failed
B
Healthy
7dpi
14dpi
30dpi
Number of cells
Mitochondrial proportion (%)
Passed
Failed
C
# of genes per cell
# of UMIs per cell
% mitochondrial genes
Healthy
7dpi
14dpi
30dpi

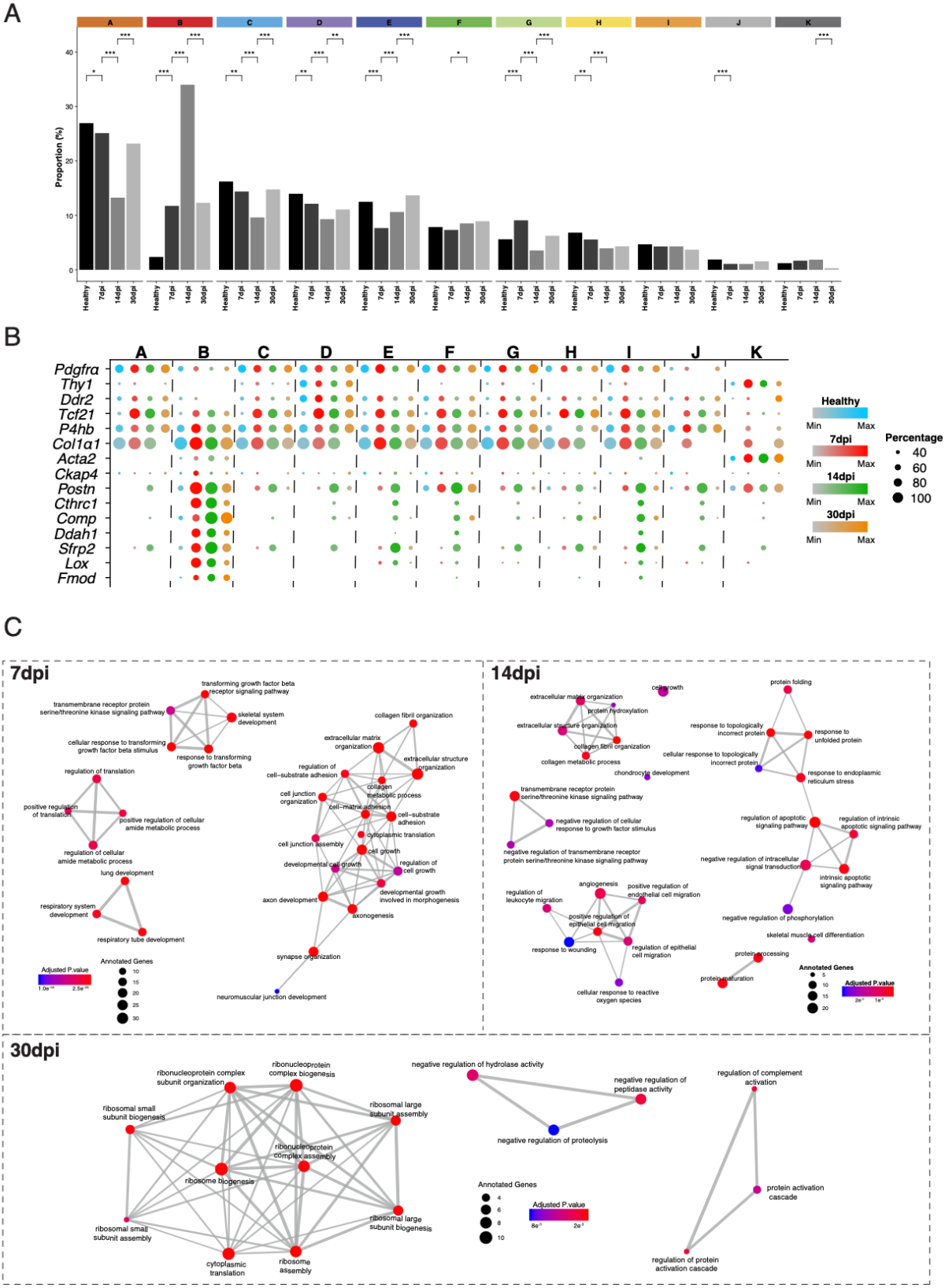
Supplementary Figure III: Quality control of single-cell RNA-sequencing analysis. **(A)** Scatter plot with number of UMIs (x-axis) and number of genes (y-axis) for the single cell analyses performed at different days after myocardial infarction. Color shows cells that passed quality filters (blue) and discarded ones (red). **(B)** Histogram of the proportion of UMIs assigned to mitochondrial genes for the single cell analyses performed at different dpi. Color shows cells that passed quality filters (blue) and discarded ones (red). **(C)** Boxplots of quality control variables (number of genes, number of UMIs and percentage of UMIs mapped to mitochondrial genes) for the different single cell datasets.


A
Pdgfra
Thy1
Ddr2
Tcf21
P4hb
Acta2
Ckap4
Sca1
Rgs5
Higd1b
Vtn
Colec11
Tinagl1
Normalized Log Expression
Cluster
A B C D E F G H I J K
B
Normalized Log Expression
Hsd11b1
Gata3
Dpep1
Ms4a4d
Nmb
Comp
Cthrc1
1500015O10Rik
Ddah1
Ptn
Pla1a
Lamb1
Efhd1
Cd248
Ly6c1
Tmem100
Ackr3
Csrl
Gm12840
Mt2
Igf1
Olfml2b
Cilp
Meox1
Cst6
Fibin
Ctgf
Slc39a1
Spry2
Notch2
Cbx3
Abhd2
Fgl2
Pdlim3
Cdc42ep3
Igfbp3
Dkk3
Ifit3
Isg15
Rtp4
Ifit1
Irf7
Gdf15
Trib3
Mthfd2
Cdkn1a
Chchd10
Rgs5
Higd1b
Vtn
Colec11
Tinagl1
C
% Cells from cluster e expressing marker - % Other cells e xpressing marker
A B C D E F G H I J K
D
Cluster
A B C D E F G H I J K
glutathione metabolic process
leukotriene metabolic process
icosanoid metabolic process
phospholipid metabolic process
regulation of cell migration
extracellular matrix organization
positive regulation of cell proliferation
negative regulation of cellular process
platelet degranulation
collagen fibril organization
glucocorticoid metabolic process
glucocorticoid biosynthetic process
organic cyclic compound catabolic process
steroid catabolic process
lipid catabolic process
positive regulation of focal adhesion assembly
positive regulation of adherens junction organization
nucleotide−sugar biosynthetic process
negative regulation of response to stimulus
positive regulation of cell junction assembly
kidney development
regulation of epithelial cell proliferation
positive regulation of epithelial cell proliferation
proteoglycan metabolic process
embryonic skeletal system development
negative regulation of transmembrane receptor protein serine/threonine kinase signaling pathway
negative regulation of cellular response to transforming growth factor beta stimulus
negative regulation of transforming growth factor beta receptor signaling pathway
regulation of transforming growth factor beta receptor signaling pathway
regulation of Ras protein signal transduction
chromatin remodeling
regulation of osteoclast development
regulation of cell projection organization
marginal zone B cell differentiation
adrenal gland development
endocrine system development
embryonic organ morphogenesis
cardiac ventricle morphogenesis
positive regulation of plasma membrane bounded cell projection assembly
cellular response to type I interferon
type I interferon signaling pathway
cytokine−mediated signaling pathway
response to cytokine
regulation of interferon−alpha production
tRNA aminoacylation
tRNA aminoacylation for protein translation
PERK−mediated unfolded protein response
response to ionizing radiation
regulation of protein serine/threonine kinase activity
muscle contraction
cell−matrix adhesion
cell junction assembly
cell−substrate adherens junction assembly
regulation of smooth muscle cell migration
Gene Ratio
0.05
0.10
0.15
0.20
0.25
EnrichR
Combined Score
1e³ 2e³ 3e³ 4e³

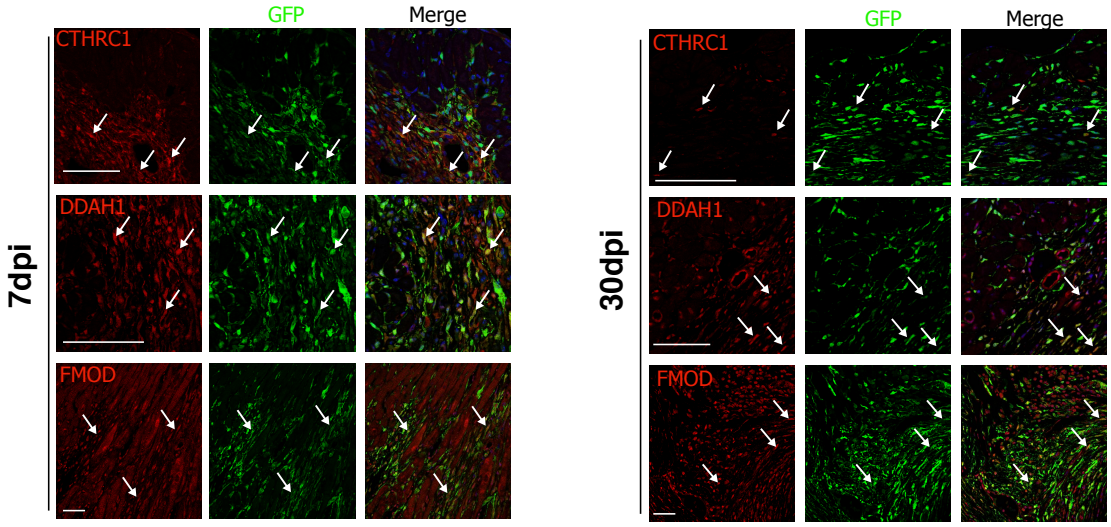
Supplementary Figure IV: All *Col1α1*-GFP^{+}/CD31^{-}/CD45^{-} clusters have specific transcriptomic signatures. (A) Violin plots showing the normalized log expression (y-axis) of traditional markers for non-activated and activated CF (*Pdgfrα, Tcf21, Acta2, Thy1, Ckap4, P4hb, Ddr2, Sca1*) per cluster (A to K) from healthy, 7, 14 and 30 dpi (left). Violin plot showing the normalized log expression (y-axis) of the five top markers in cluster K for pericytes (*Rgs5, Higd1b, Vtn, Tinagl1, Colec11,*) per cluster (A to K) from healthy, 7, 14 and 30 dpi (right). **(B)** Violin plots of the normalized log expression of the top five markers per cluster. Color indicates the corresponding cluster (A to K). **(C)** Boxplot denoting the marker specificity of every identified cluster. The specific (y-axis) value corresponds to the percentage of cells of a particular cluster expressing a given marker minus the percentage of every other cell expressing the same marker. **(D)** Dot plot comparison of enriched GO terms for each of the identified clusters based on the expression of marker genes per cluster.


A
Col1α1-GFP
Farbehi et al
UMAP 1
UMAP 2
B
Clustering Farbehi et al, 2019
Col1a1 (Fibroblast)
Postn (Myofibroblast)
Plp1 (Glial)
Kdr (Endothelial)
Cd79a (B-cells)
Cd68 (Macrophage)
Vtn (Mural)
Ccl5 (NK cells)
Cd3d (T-cells)
Ccnb2 (Proliferating)
C
Clustering GFP+ CF
Cluster A
Cluster B
Cluster C
Cluster D
Cluster E
Cluster F
Cluster G
Cluster H
Cluster I
Cluster J
Cluster K

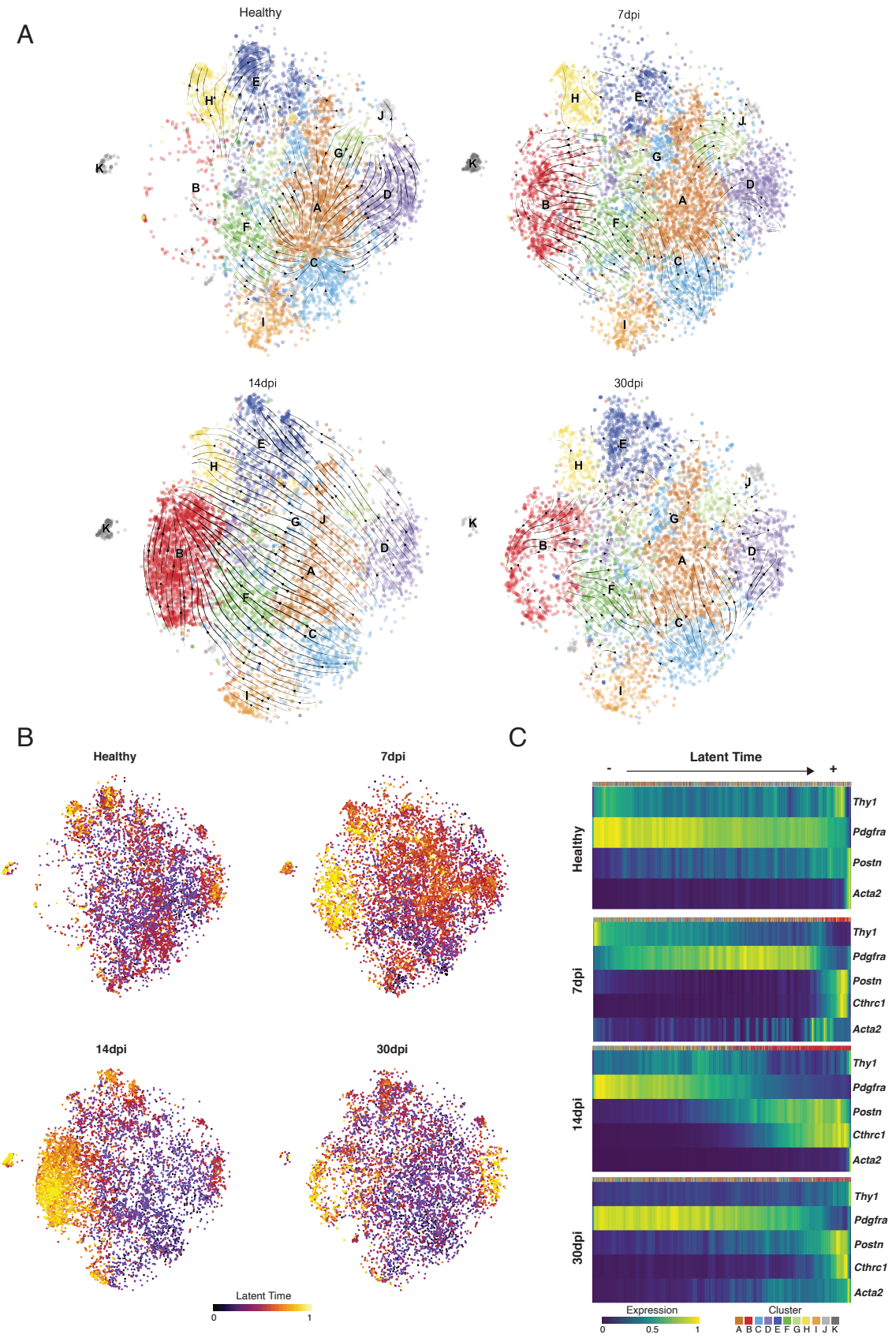
Supplementary Figure V: *Col1α1*-GFP^{+}/CD31^{-}/CD45^{-} heterogeneity reveals no differences with a recently published unbiased model. (A) UMAP representation of *Col1α1*-GFP^{+}/CD31^{-}/CD45^{-} heterogeneity (red dots) on top of Farbehi *et al*. (2019) data (blue dots) for total cardiac interstitial cells (CIC) population at 7dpi (left). UMAP representation of total CIC population from Farbehi *et al*. clusters through unsupervised analysis (right). **(B)** Individual UMAP representation of the top marker genes that define each cluster in Farbehi *et al*. **(C)** Individual UMAP representation for the score of the transcriptomic signature of each Col1α1-GFP^{+}/CD31^{-}/CD45^{-} cluster on top of the single-cell data obtained from the total CIC population.


A
A
B
C
D
E
F
G
H
I
J
K
Proportion (%)
Healthy
7dpi
14dpi
30dpi
B
Pdgfra
Thy1
Ddr2
Tcf21
P4hb
Col1a1
Acta2
Ckap4
Postn
Cthrc1
Comp
Ddah1
Sfrp2
Lox
Fmod
Healthy
7dpi
14dpi
30dpi
Min
Max
Percentage
40
60
80
100
C
7dpi
transmembrane receptor protein serine/threonine kinase signaling pathway
transforming growth factor beta receptor signaling pathway
skeletal system development
cellular response to transforming growth factor beta stimulus
response to transforming growth factor beta
regulation of translation
positive regulation of translation
positive regulation of cellular amide metabolic process
regulation of cellular amide metabolic process
lung development
respiratory system development
respiratory tube development
collagen fibril organization
extracellular matrix organization
extracellular structure organization
regulation of cell–substrate adhesion
collagen metabolic process
cell junction organization
cell–matrix adhesion
cell–substrate adhesion
cytoplasmic translation
cell junction assembly
cell growth
developmental cell growth
regulation of cell growth
developmental growth involved in morphogenesis
axon development
axonogenesis
synapse organization
neuromuscular junction development
Adjusted P.value
Annotated Genes
14dpi
cell growth
extracellular matrix organization
protein hydroxylation
extracellular structure organization
collagen fibril organization
collagen metabolic process
chondrocyte development
protein folding
response to topologically incorrect protein
response to unfolded protein
cellular response to topologically incorrect protein
response to endoplasmic reticulum stress
transmembrane receptor protein serine/threonine kinase signaling pathway
negative regulation of cellular response to growth factor stimulus
negative regulation of transmembrane receptor protein serine/threonine kinase signaling pathway
regulation of apoptotic signaling pathway
regulation of intrinsic apoptotic signaling pathway
negative regulation of intracellular signal transduction
intrinsic apoptotic signaling pathway
angiogenesis
positive regulation of endothelial cell migration
regulation of leukocyte migration
positive regulation of epithelial cell migration
response to wounding
regulation of epithelial cell migration
negative regulation of phosphorylation
skeletal muscle cell differentiation
protein processing
protein maturation
cellular response to reactive oxygen species
Annotated Genes
Adjusted P.value
30dpi
ribonucleoprotein complex subunit organization
ribonucleoprotein complex biogenesis
ribosomal small subunit biogenesis
ribosomal large subunit assembly
ribonucleoprotein complex assembly
ribosome biogenesis
ribosomal small subunit assembly
ribosomal large subunit biogenesis
cytoplasmic translation
ribosome assembly
negative regulation of hydrolase activity
negative regulation of peptidase activity
negative regulation of proteolysis
regulation of complement activation
protein activation cascade
regulation of protein activation cascade
Annotated Genes
Adjusted P.value

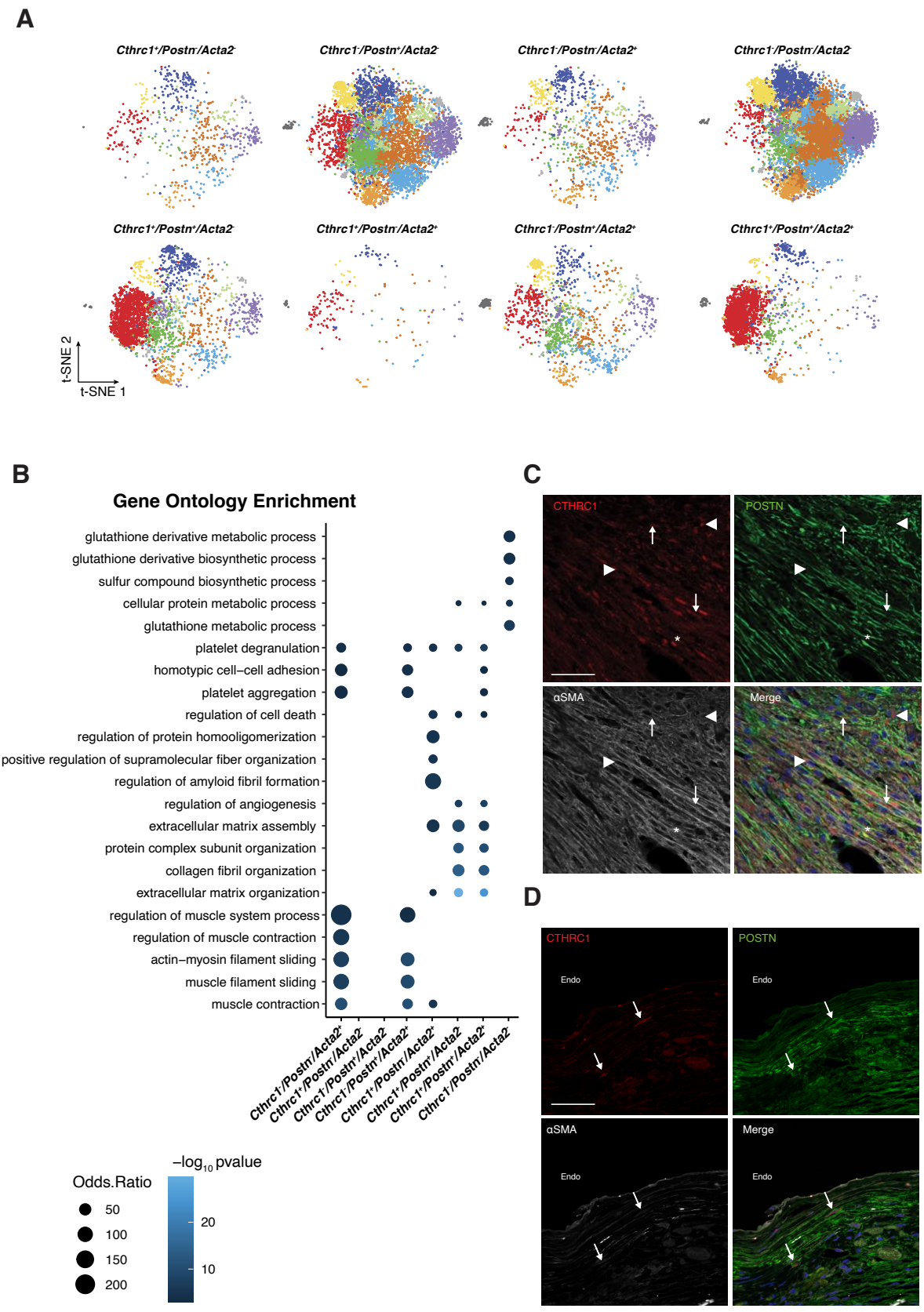
Supplementary Figure VI: The dynamics and specificity of the cluster B cells transcriptomic signature indicate a role in the healing process. (A) Bar plots showing the percentage of cells per cluster (y-axis) relative to the total number of cells at each time point (grey scale). Proportion test. * p≤0.05, ** p≤0.01, *** p≤0.001. **(B)** Dot plot comparison of expression and specificity of markers for non-activated fibroblasts (*Pdgfrα, Thy1, Ddr2, Tcf21, P4hb, Col1α1*), activated fibroblasts (*Acta2, Ckap4, Postn*) and cluster B markers (*Cthrc1, Comp, Ddah1, Sfrp2, Lox, Fmod*) in healthy myocardium and at 7, 14 and 30 dpi. Dot size represents percentage of cells, included in the cluster that express a given marker, and the expression level defined by color intensity. **(C)** Network representation of enriched pathways for the defined cluster B cells at different timepoints after MI. Dot size represents the number of cluster B markers annotated for each pathway and color the statistical significance.



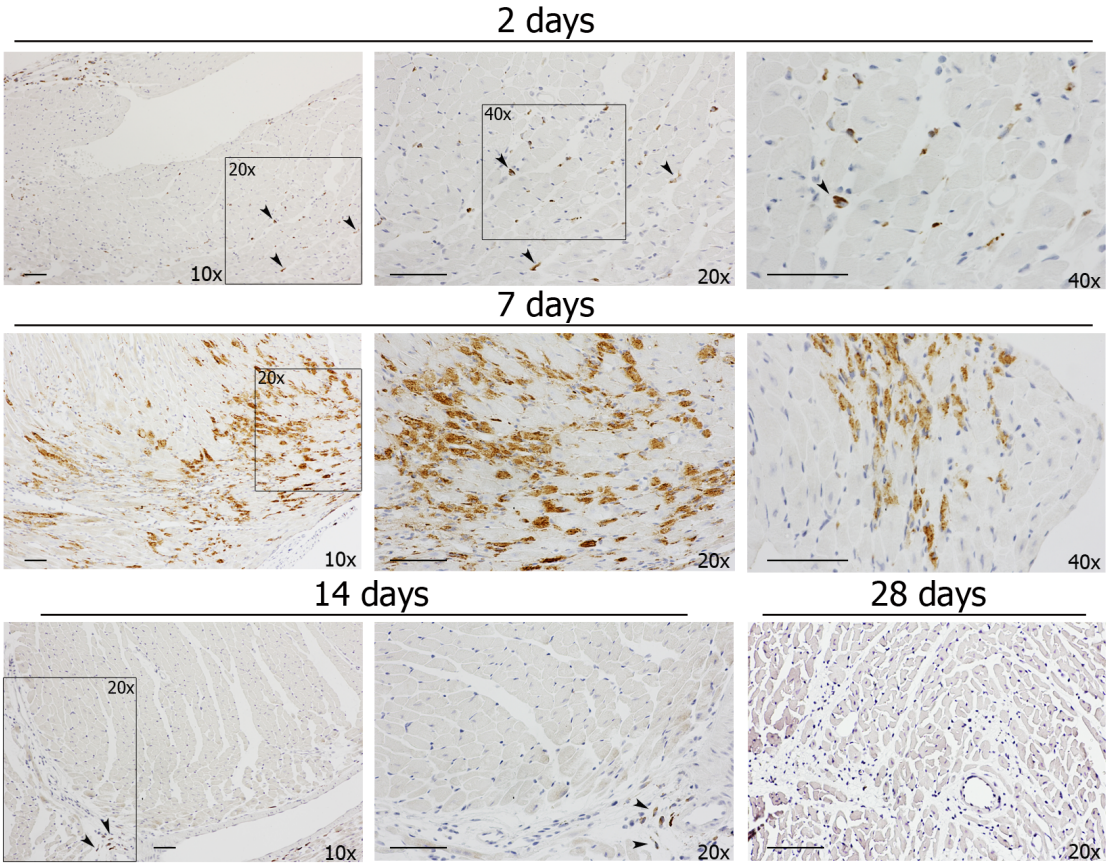
Supplementary Figure VII: Cluster B is specifically located in the infarct zone along MI. Spatial location of CTHRC1, DDAH1, or FMOD in the infarct zone at 7 and 30 dpi. GFP^+ (green), CTHRC1, DDAH1, or FMOD (red), Nuclei (DAPI, blue). Co-localizations are in yellow (arrows). Scale bars: 50 µm.


A
Healthy
7dpi
14dpi
30dpi
B
Healthy
7dpi
14dpi
30dpi
Latent Time
0
1
C
Latent Time
-
+
Healthy
7dpi
14dpi
30dpi
Thy1
Pdgfra
Postn
Cthrc1
Acta2
Expression
0
0.5
1
Cluster
A B C D E F G H I J K

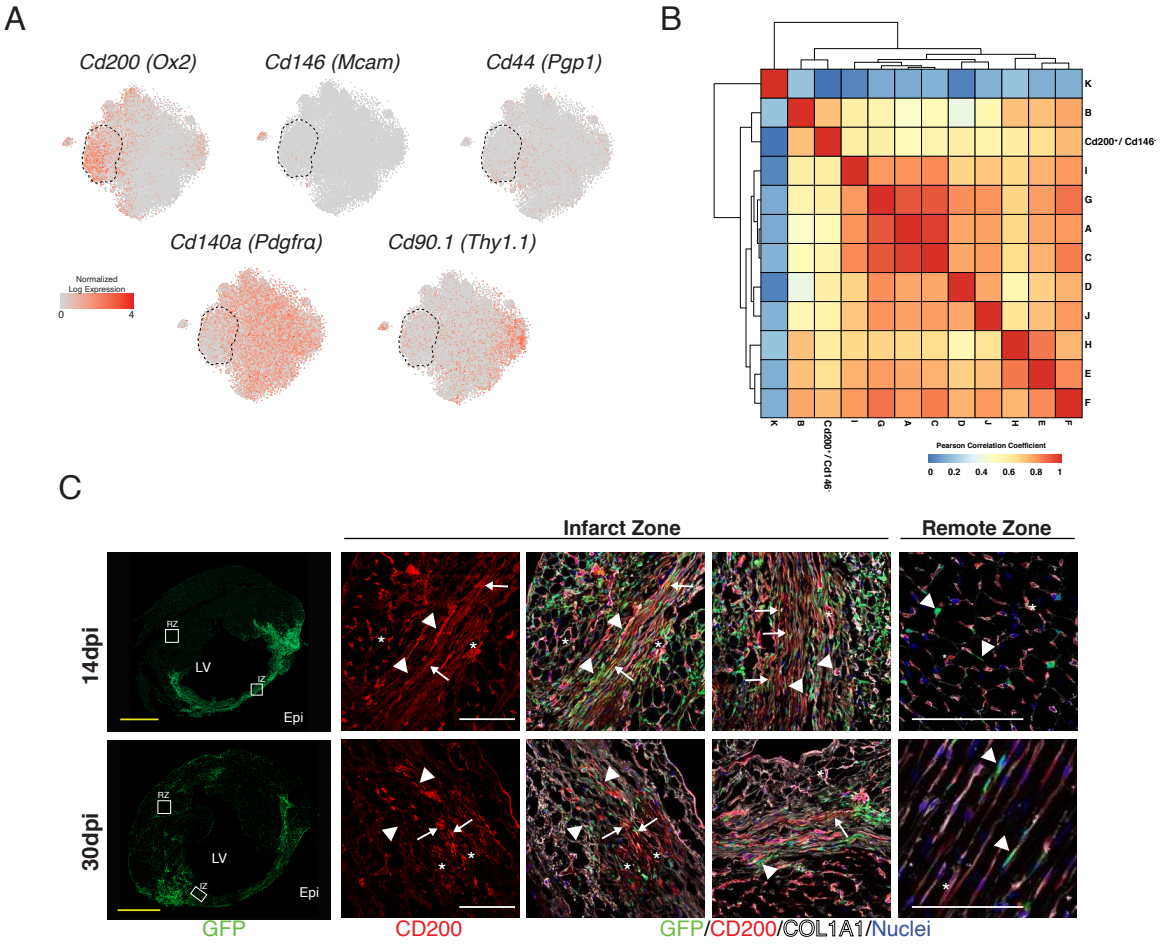
Supplementary Figure VIII: Trajectory analysis reveals dynamic transitions between clusters and cluster B cells as a transition cluster with *Cthrc1* expression specific to the final stage of activation. (A) t-SNE representation of the RNA velocity results of the single cell data sets at different time points. **(B)** t-SNE plots colored by the recovered latent time in the single cell data sets at different time points. **(C)** Expression heatmaps for different genes (*Thy1*, *Pdgfrα*, *Postn*, *Cthrc1* and *Acta2*) with CF ordered by latent time values (left – and right +) at different time points. Cluster identity is represented in the top.


A
Cthrc1+/Postn-/Acta2-
Cthrc1-/Postn+/Acta2-
Cthrc1-/Postn-/Acta2+
Cthrc1-/Postn-/Acta2-
Cthrc1+/Postn+/Acta2-
Cthrc1+/Postn-/Acta2+
Cthrc1-/Postn+/Acta2+
Cthrc1+/Postn+/Acta2+
t-SNE 2
t-SNE 1
B
Gene Ontology Enrichment
glutathione derivative metabolic process
glutathione derivative biosynthetic process
sulfur compound biosynthetic process
cellular protein metabolic process
glutathione metabolic process
platelet degranulation
homotypic cell–cell adhesion
platelet aggregation
regulation of cell death
regulation of protein homooligomerization
positive regulation of supramolecular fiber organization
regulation of amyloid fibril formation
regulation of angiogenesis
extracellular matrix assembly
protein complex subunit organization
collagen fibril organization
extracellular matrix organization
regulation of muscle system process
regulation of muscle contraction
actin–myosin filament sliding
muscle filament sliding
muscle contraction
Cthrc1-/Postn-/Acta2-
Cthrc1+/Postn-/Acta2-
Cthrc1-/Postn+/Acta2-
Cthrc1-/Postn-/Acta2+
Cthrc1+/Postn+/Acta2-
Cthrc1+/Postn-/Acta2+
Cthrc1-/Postn+/Acta2+
Cthrc1+/Postn+/Acta2+
−log10 pvalue
Odds.Ratio
50
100
150
200
20
10
C
CTHRC1
POSTN
αSMA
Merge
D
CTHRC1
POSTN
αSMA
Merge
Endo

Supplementary Figure IX: Cluster B cells are a subpopulation of activated *Postn*$^{+}$ CF. (A) t-SNE plots of different CF subgroups based on the expression of *Postn*, *Acta2* and *Cthrc1* and color coded according to clusters. **(B)** Dot plot comparison of enriched GO terms for each of the subgroups. Dot size represents the odds ratio and the color denotes statistical significance. **(C)** Immunohistochemistry of CTHRC1 (red), POSTN (green) and αSMA (white) at 7dpi and **(D)** at 60dpi (DAPI/nuclei, blue). CTHRC1^{+}/POSTN^{-}/αSMA^{-} (arrowheads), CTHRC1^{+}/POSTN^{+}/αSMA^{-} (arrows) and CTHRC1^{+}/POSTN^{+}/αSMA^{+} (asterisk). Scale bars: 50 µm. Abbreviations as in Supplementary Figure I.

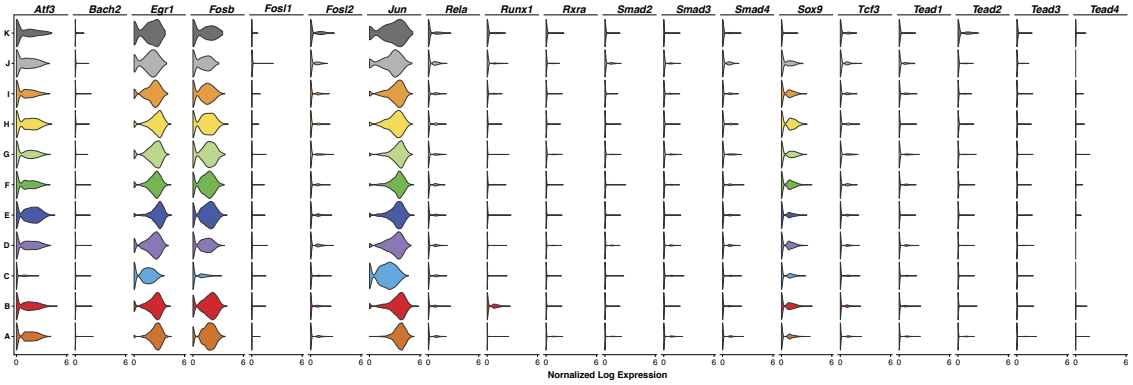


Supplementary Figure X: CTHRC1^{+} CF are also present in a chronic cardiac fibrosis model. Representative images of the immunohistochemistry analysis of CTHRC1^{+} CF (in brown color, arrowheads) performed on sections of the left ventricle of murine hearts 2, 7, 14 and 28 days after the treatment with angiotensin-II. Scale bars: 50 µm.

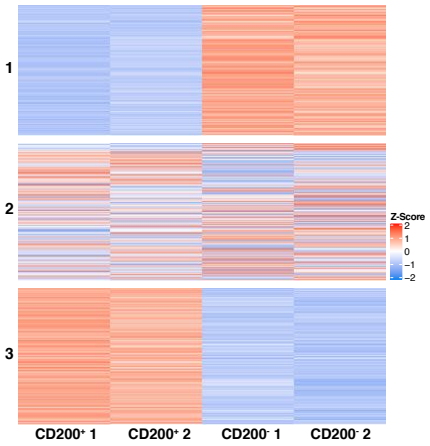

A
Cd200 (Ox2)
Cd146 (Mcam)
Cd44 (Pgp1)
Cd140a (Pdgfra)
Cd90.1 (Thy1.1)
Normalized Log Expression
0
4
B
K
B
Cd200+/ Cd146+
I
G
A
C
D
J
H
E
F
Pearson Correlation Coefficient
0 0.2 0.4 0.6 0.8 1
C
Infarct Zone
Remote Zone
14dpi
30dpi
RZ
LV
IZ
Epi
GFP
CD200
GFP/CD200/COL1A1/Nuclei
D
Cd200
Mcam
Cd200/Mcam
Expression
10.0
7.5
5.0
2.5
2
4
6
8
10
tSNE 2
tSNE 1
1402 cells
96 cells
85 cells
Cluster
A B C D E F G H I J K
Proportion (%)
60
40
20
0
7 37 4 16 7 10 3 9 3 1 2
9 27 1 4 5 3 2 4 0 3 41
0 33 1 1 1 2 0 1 1 1 58

Supplementary Figure XI: CD200$^+$ CF are located in the scar region and are enriched in RCF. (A) t-SNE representation showing expression of *Cd200 (Ox2), Cd146 (Mcam), Cd44 (Pgp1), Cd140a (Pdgfrα),* and *Cd90.1 (Thy1.1)* in the global scRNA-seq. **(B)** Correlation heatmap of RCF markers between scRNA-seq aggregated individual cluster expression and bulk RNA-seq expression in GFP$^+$/ CD200$^+$/CD146$^-$/CD31$^-$/CD45$^-$ cells isolated at 7dpi. **(C)** Transverse sections of *Col1α1-GFP* hearts at 14 and 30 dpi. Scale bars: 1 mm. Immunofluorescence analysis of GFP$^+$ (green), CD200$^+$ (red), COL1α1$^+$ (grey) and DAPI/nuclei (blue) at 14 and 30 dpi. Scale bars: 100 μm. Co-localization of GFP$^+$ and CD200$^+$ in yellow (arrows) in IZ (and co-localize with COL1α1 in light yellow). Arrowheads indicate GFP$^+$/CD200$^-$ cells and asterisks indicate GFP$^-$/CD200$^+$. **(D)** t-SNE plots with the expression of the membrane markers *Cd200*, *Mcam (Cd146) or both* from the scRNA-seq at 7dpi. The heatmap denotes the colors to identify cells expressing both markers (top row). t-SNE plots with cells colored by cluster. The number denotes the cells identified as *Cd200$^+$/Cd146$^-$, Cd200$^-$/Cd146$^-$* and *Cd200$^+$/Cd146$^+$* (from left to right, middle row). Bar plot showing the proportion of *Cd200$^+$/Cd146$^-$, Cd200$^-$/Cd146$^-$* and *Cd200$^+$/Cd146$^+$*cells per cluster in the corresponding column (bottom row). Abbreviations as in Supplementary Figure I.

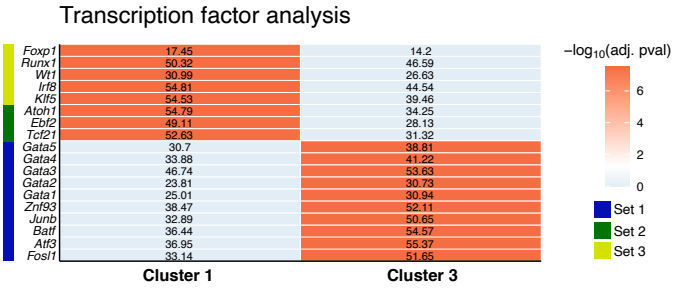
A


Atf3
Bach2
Egr1
Fosb
Fosl1
Fosl2
Jun
Rela
Runx1
Rxra
Smad2
Smad3
Smad4
Sox9
Tcf3
Tead1
Tead2
Tead3
Tead4
K
J
I
H
G
F
E
D
C
B
A
Nornalized Log Expression

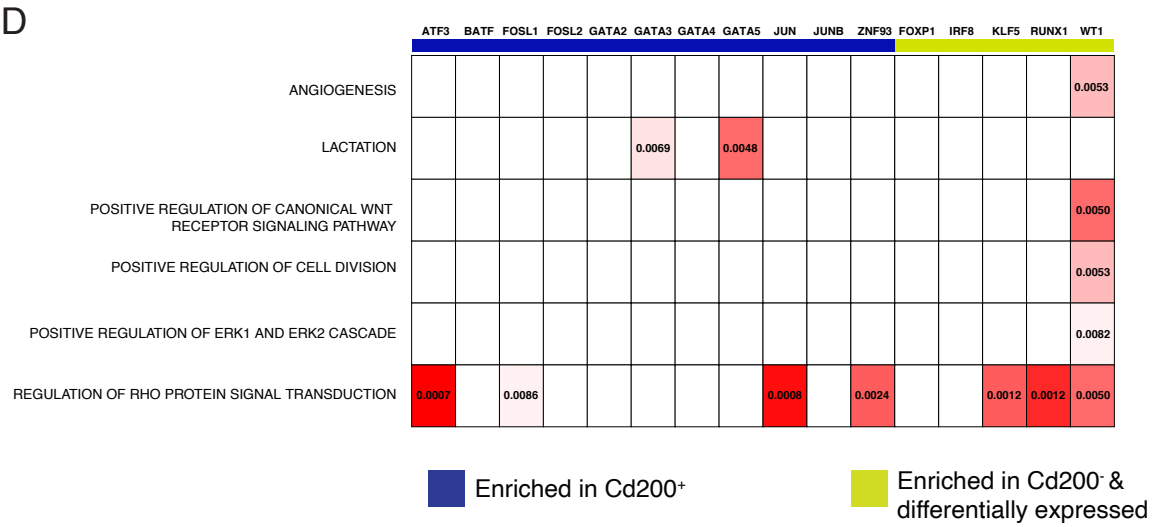
B


1
2
3
Z-Score
2
1
0
−1
−2
CD200+ 1
CD200+ 2
CD200- 1
CD200- 2

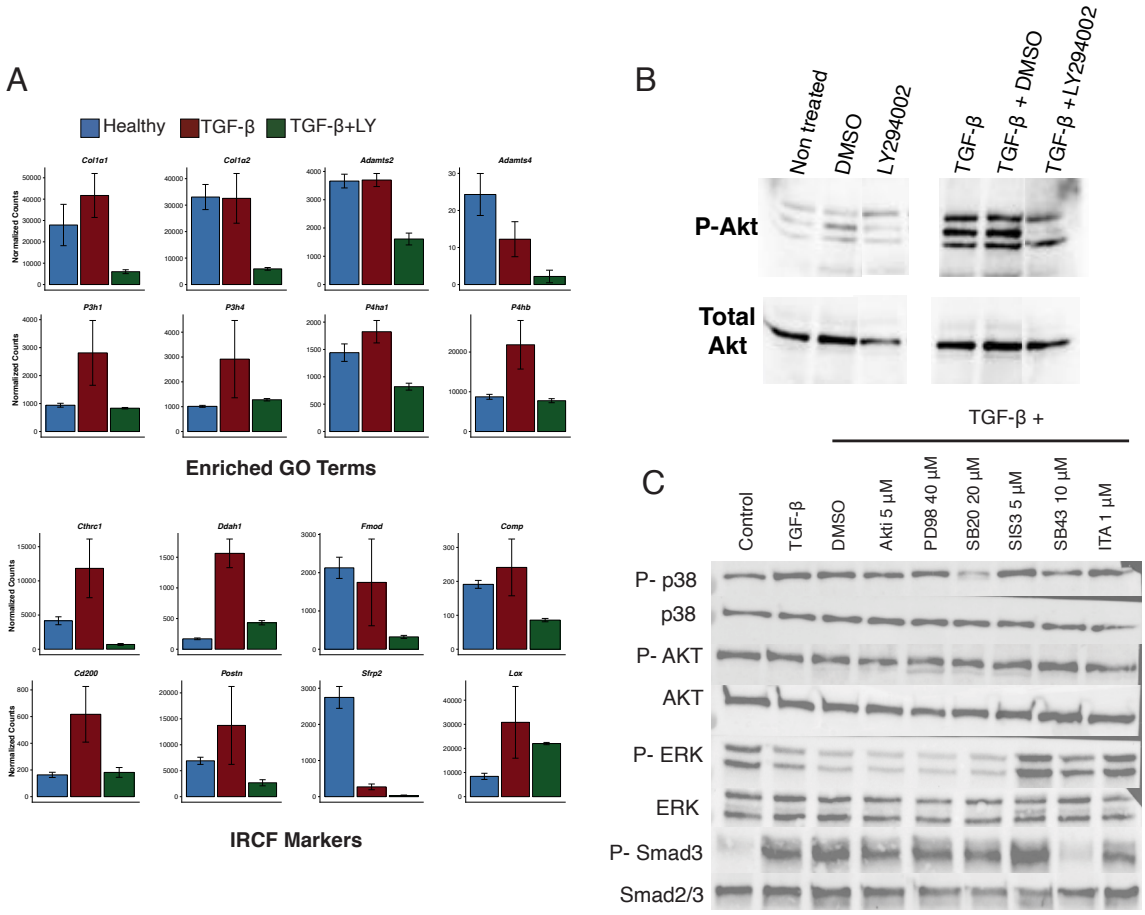
C


Transcription factor analysis
Foxp1 17.45 14.2
Runx1 50.32 46.59
Wt1 30.99 26.63
Irf8 54.81 44.54
Klf5 54.53 39.46
Atoh1 54.79 34.25
Ebf2 49.11 28.13
Tcf21 52.63 31.32
Gata5 30.7 38.81
Gata4 33.88 41.22
Gata3 46.74 53.63
Gata2 23.81 30.73
Gata1 25.01 30.94
Znf93 38.47 52.11
Junb 32.89 50.65
Batf 36.44 54.57
Atf3 36.95 55.37
Fosl1 33.14 51.65
Cluster 1
Cluster 3
−log10(adj. pval)
6
4
2
0
Set 1
Set 2
Set 3

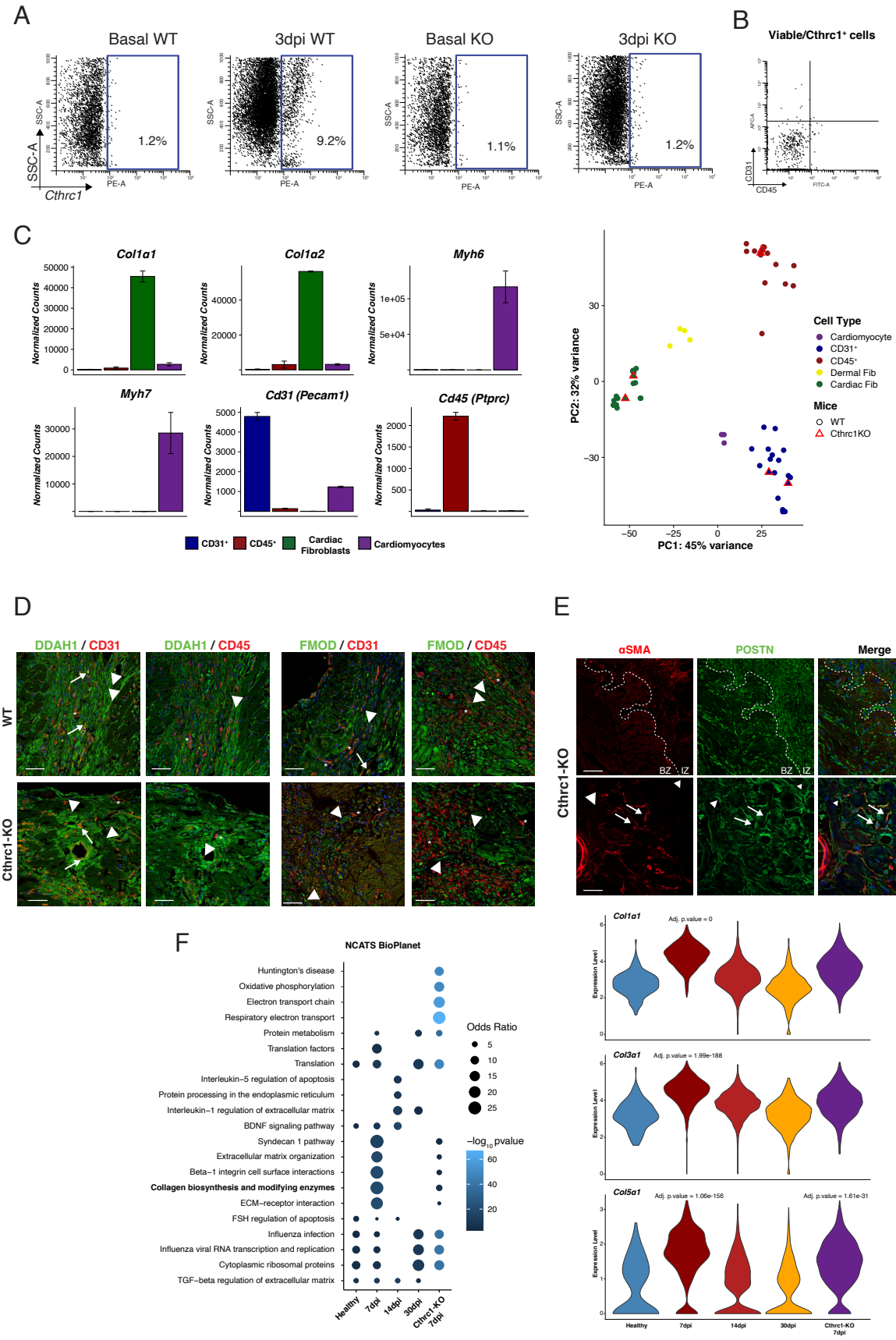
D


ATF3 BATF FOSL1 FOSL2 GATA2 GATA3 GATA4 GATA5 JUN JUNB ZNF93 FOXP1 IRF8 KLF5 RUNX1 WT1
ANGIOGENESIS
LACTATION
POSITIVE REGULATION OF CANONICAL WNT RECEPTOR SIGNALING PATHWAY
POSITIVE REGULATION OF CELL DIVISION
POSITIVE REGULATION OF ERK1 AND ERK2 CASCADE
REGULATION OF RHO PROTEIN SIGNAL TRANSDUCTION
0.0053
0.0069
0.0048
0.0050
0.0053
0.0082
0.0007
0.0086
0.0008
0.0024
0.0012
0.0012
0.0050
Enriched in Cd200+
Enriched in Cd200- & differentially expressed

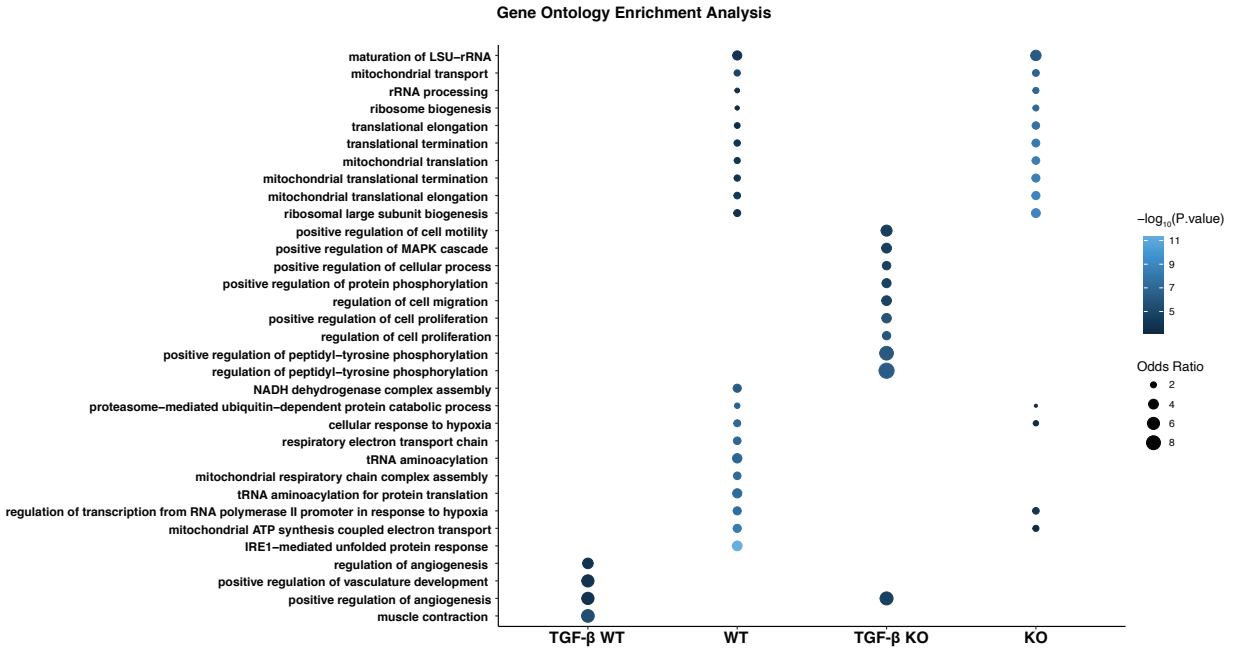
Supplementary Figure XII: Different genomic approaches reveal high complexity in the regulation of RCF. (A) Violin plot representation of the expression of TF (*Atf3, Bach2, Egr1, Fosb, Fosl1, Fosl2, Jun, Rela, Runx1, Rxra, Smad2, Smad3, Smad4, Sox9, Tcf3, Tead1, Tead2, Tead3, Tead4*) in all the clusters. **(B)** Heatmap of the clustering of the peaks from the ATAC-seq analysis showing the accessibility of the peaks (rows) in the four samples (columns). Red denotes highly and blue poorly accessible sites. Cluster 1 (Cluster 3) is enriched in peaks accessible in $CD200^-$ ($CD200^+$) cells.**(C)** Plot of the TF enrichment analysis in Cluster 1 (left column, $CD200^-$) and Cluster 2 (right column, $CD200^+$). Color denotes the –log10 (p-value) associated with the Fisher test analysis of enrichment. TF are ranked based on the different % of accessibility (shown in boxes): from more decreased accessibility of $CD200^-$ to $CD200^+$ (top) to more increased accessibility (bottom). **(D)** Summary of pathways obtained by gene-set analysis for genes associated with enriched peaks in Set 1 (blue) and Set 4 (yellow).


A
Healthy
TGF-β
TGF-β+LY
Col1a1
Col1a2
Adamts2
Adamts4
P3h1
P3h4
P4ha1
P4hb
Normalized Counts
Enriched GO Terms
Cthrc1
Ddah1
Fmod
Comp
Cd200
Postn
Sfrp2
Lox
IRCF Markers
B
Non treated
DMSO
LY294002
TGF-β
TGF-β + DMSO
TGF-β +LY294002
P-Akt
Total Akt
C
TGF-β +
Control
TGF-β
DMSO
Akti 5 μM
PD98 40 μM
SB20 20 μM
SIS3 5 μM
SB43 10 μM
ITA 1 μM
P- p38
p38
P- AKT
AKT
P- ERK
ERK
P- Smad3
Smad2/3
D
SB-431542
SIS3
PD-98059
Akti
SB-205380
ITA
Phenotypic transformation
Pro-fibrotic characteristics
Relative quantification (Nc)
Cthrc1
Ddah1
Cluster B markers
Postn
Acta2
Activated Fibroblast
Myofibroblast
Col1a1
Col3a1
Lox

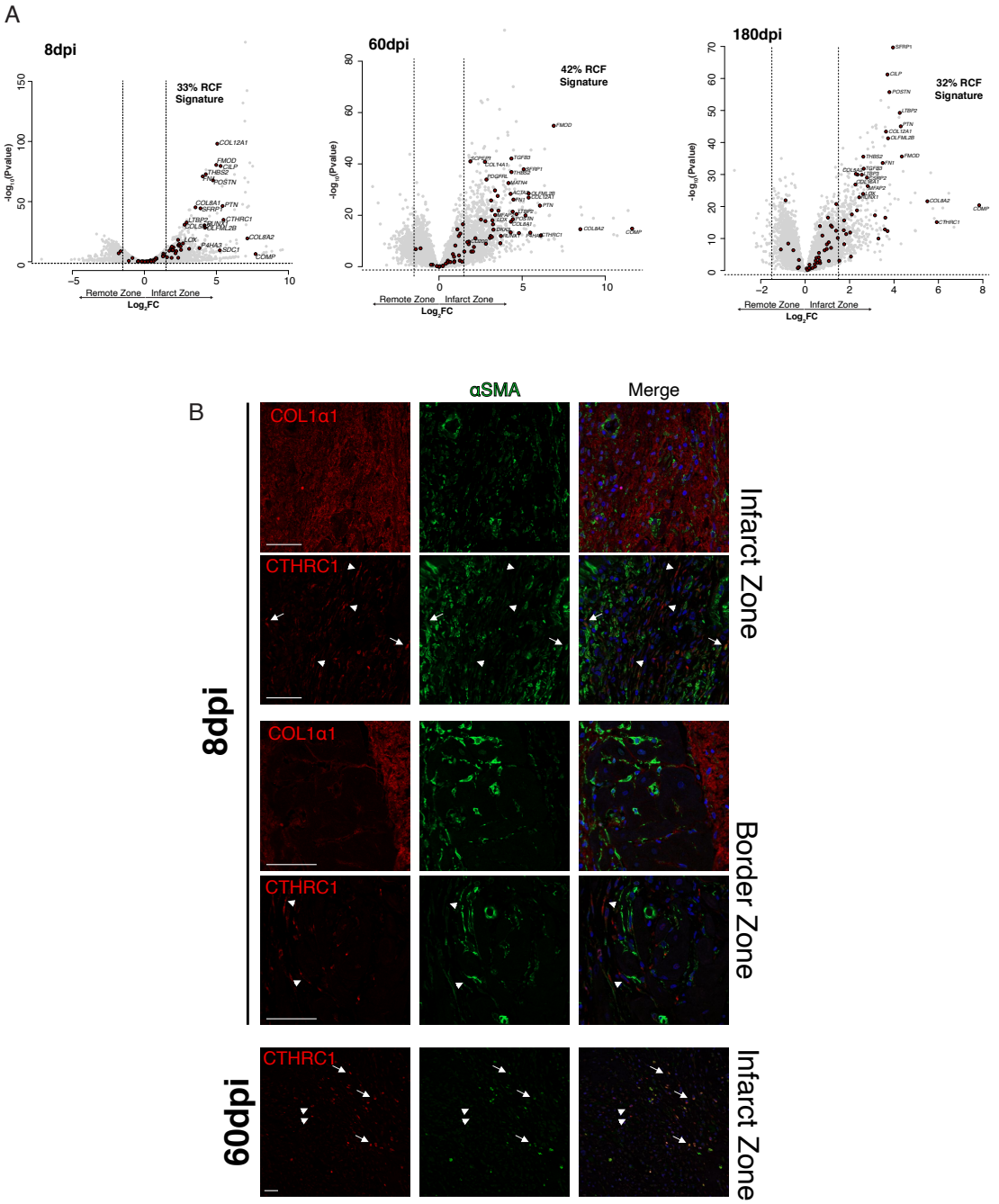
Supplementary Figure XIII: The non-canonical TGF-β signaling pathway has a role in the regulation of RCF transcriptomic signature. (A) Normalized expression bar plots (mean ± SD) for genes annotated to enriched GO terms (above) and top RCF markers (below). **(B)** Western blot for p-Akt and total Akt protein in cell lysates from CF cultured for 1 hour without treatment, with DMSO, LY294002, TGF-β, TGF-β + DMSO and TGF-β + LY294002. **(C)** Western blots showing the effect of small molecules in the canonical and non-canonical TGF-β signaling pathway. **(D)** Boxplots showing the level of expression of genes that define a phenotype transformation to RCF and a pro-fibrotic profiling related to the extracellular matrix organization.


A
Basal WT
3dpi WT
Basal KO
3dpi KO
1.2%
9.2%
1.1%
1.2%
SSC-A
PE-A
Cthrc1
B
Viable/Cthrc1⁺ cells
CD31
CD45
C
Col1a1
Col1a2
Myh6
Myh7
Cd31 (Pecam1)
Cd45 (Ptprc)
Normalized Counts
CD31⁺
CD45⁺
Cardiac Fibroblasts
Cardiomyocytes
PC2: 32% variance
PC1: 45% variance
Cell Type
Cardiomyocyte
CD31⁺
CD45⁺
Dermal Fib
Cardiac Fib
Mice
WT
Cthrc1KO
D
DDAH1 / CD31
DDAH1 / CD45
FMOD / CD31
FMOD / CD45
WT
Cthrc1-KO
E
αSMA
POSTN
Merge
BZ
IZ
F
NCATS BioPlanet
Huntington's disease
Oxidative phosphorylation
Electron transport chain
Respiratory electron transport
Protein metabolism
Translation factors
Translation
Interleukin−5 regulation of apoptosis
Protein processing in the endoplasmic reticulum
Interleukin−1 regulation of extracellular matrix
BDNF signaling pathway
Syndecan 1 pathway
Extracellular matrix organization
Beta−1 integrin cell surface interactions
Collagen biosynthesis and modifying enzymes
ECM−receptor interaction
FSH regulation of apoptosis
Influenza infection
Influenza viral RNA transcription and replication
Cytoplasmic ribosomal proteins
TGF−beta regulation of extracellular matrix
Healthy
7dpi
14dpi
30dpi
Cthrc1-KO 7dpi
Odds Ratio
−log10pvalue
Col1a1
Col3a1
Col5a1
Adj. p.value = 0
Adj. p.value = 1.99e-188
Adj. p.value = 1.06e-156
Adj. p.value = 1.61e-31
Expression Level

Supplementary Figure XIV: The absence of *Cthrc1* exclusively affects the function of CF. (A) Flow cytometry analysis of CTHRC1^{+} CIC in WT and KO mice in healthy myocardium (right) and at 3dpi (left). **(B)** CD31 and CD45 expression in viable, CTHRC1^{+}/CD31^{-}/CD45^{-} CIC in WT mice at 3dpi. **(C)** Normalized expression bar plots (mean ± SD) of expression levels for specific markers of cell types (*Col1α1, Col1α2, Myh6, Myh7 Cd31* (*Pecam1*), *Cd45* (*Ptprc*)) in different cell fractions (CD31^{+} endothelial cells, CD45^{+} bone marrow-derived cells, CF, cardiomyocytes) in WT mice. PCA scatter plot of cardiomyocytes, endothelial (CD31) and bone marrow-derived (CD45) cells, dermal fibroblasts and CF from *Cthrc1-KO* (triangles) and WT (squares) mice subjected to transcriptome profiling. **(D)** Spatial location of DDAH1, FMOD, CD31 and CD45 in the IZ of WT and *Cthrc1-KO* mice at 5dpi. DDAH1 or FMOD (green, arrowheads), CD31 or CD45 (red, asterisks), Nuclei (DAPI, blue). Co-localizations are in yellow (arrows). Scale bars: 100 µm. **(E)** Immunohistochemistry of αSMA (red) and POSTN (green) in *Cthrc1-KO* mice at 5dpi (DAPI/nuclei, blue). Representative images of POSTN^{+} (arrowheads, green) and POSTN^{+}/αSMA^{+} (arrows, yellow) in the border zone. Dotted lines delimit IZ from BZ. Scale bars: 50 µm. **(F)** Dot plot with the enriched signaling pathways for CF in each of the conditions under study. Dot size represents the odds ratio and the color denotes statistical significance (left). Violin plots showing the expression levels of different collagen mRNA molecules (*Col1α1*, *Col3α1* and *Col5α1*) in RCF populations identified in each of the single cell datasets. Wilcoxon test. (right). CIC = cardiac interstitial cells; IZ = infarct zone; KO = knockout; WT = wild type.



Supplementary Figure XV: *Cthrc1-KO* CF have a blunted response to TGF-β compared to WT CF. Dot plot of enriched GO categories in bulk RNA-seq profiles of cultured WT and *Cthrc1-KO* CF. Dot size represents the odds ratio and the color denotes statistical significance.



Supplementary Figure XVI: Dynamics of RCF along the time course of myocardial infarction in pigs. (A) Volcano plots show differentially expressed genes between infarct and remote zones (left). Red dots indicate porcine transcripts homologous to RCF markers. Differential gene expression analysis detects 33% (8dpi), 42% (60dpi) and 32% (180dpi) of the RCF signature in the IZ. **(B)** Immunohistochemistry of CTHRC1, COL1α1 and αSMA of infarct and border zones of infarcted swine heart at 8 and 60 dpi. $CTHRC1^+$ (arrowheads) and $CTHRC1^+/\alpha SMA^+$ CF (arrows). Scale bars: 50 µm.